Panel Summary Report
of the
Groundfish Assessment Review Meeting
(GARM III)

Part 2. Assessment Methodology (Models)

By

Chairman: Robert O'Boyle

Review Panel: José De Oliveira, Stratis Gavaris, Jim Ianelli, Yan Jiao, Cynthia Jones and Paul Medley

Report Date: March 25, 2008

Meeting Dates and Location:
25 – 29 February 2008
Northeast Fisheries Science Center
Stephen H. Clark Conference Room
Woods Hole, Massachusetts

TABLE OF CONTENTS

SUMMARY

Models to assess the status and productivity characteristics of 19 New England groundfish stocks being considered at the 2008 Groundfish Review Assessment Meetings (GARM) were reviewed at the Northeast Fisheries Science Center in Woods Hole, Massachusetts during 25 – 29 February 2008. The review considered the applicability of a number of modeling approaches (relative trends to age and length – based models) and examined the utility of statistical catch-at-age formulations to address specific issues on each stock. The latter included the ability of these models to address observed retrospective patterns, the potential causes of which were also considered at the meeting. The review focused on the key observational error and model process assumptions for each stock to determine which class of assessment model was most appropriate to use and provided guidance on model formulations to address specific issues for each stock. The sufficiency of the assessment models to provide Biological Reference Points was also discussed as were the implications of zeros in fishery independent indices.

This was the second meeting of a four part process, with the first on data inputs (29 October – 2 November 2007) and the remaining two on biological reference points (28 April – 2 May 2008) and review of the assessments (4 – 8 August 2008). The GARM process has been designed so that each review can inform subsequent ones.

The body of this report consists of the recommendations of a seven member review panel in response to the meeting’s terms of reference. The report also includes a synopsis of each of the working papers presented at the meeting along with the associated discussion, during which suggestions and recommendations were made to address identified issues. The panel considered these in drafting this report.

Overall, the meeting successfully fulfilled its terms of reference and represents an important contribution to the GARM III process.

INTRODUCTION

This document is the summary report of the review panel (herein termed the 'Panel') of the Groundfish Assessment Review Meeting (GARM) on assessment methodology. The GARM is a regional scientific peer review process developed in 2002 to provide assessments for the stocks managed under the Northeast Multispecies Fishery Management Plan (Multispecies FMP). The first two GARMs took place in October 2002 (NEFSC, 2002a) and August 2005 (NEFSC, 2005) respectively. This GARM III is the most comprehensive to date, intended to provide peer reviewed assessments on 19 groundfish stocks managed by the New England Fisheries Management Council (NEFMC).

The four meetings of GARM III include:

- Data Inputs (29 Oct – 2 Nov 2007)
- Assessment Methodology (25 – 29 Feb 2008)
- Biological Reference Points (28 April – 2 May 2008)
- Assessments (4 – 8 August 2008)

The first three meetings are to establish the analytical formulations of the assessments to be used in the last meeting. The first meeting (NEFSC, 2007) focused on the data inputs (e.g. catch, sampling, surveys, etc) to be used in the assessments. The second meeting, which is the focus of this report, considered the assessment approaches to be applied to the datasets of each stock discussed at the first meeting (see Terms of Reference, appendix 1). The applicability of classes of assessment models, from relative trends to age and length-based models, was considered as well as the ability of the age-structured models to address observed retrospective patterns. Additional issues included treatment of zeros in the evaluation of fishery independent indices (e.g. surveys) and the capacity of assessment models to provide measures of stock status consistent with biological reference points (BRPs). The latter will be an important consideration for the third meeting, which will focus on the determination of the biological reference points used to guide management decision – making (see terms of reference, appendix 2).

Many of the assessments displayed similar issues, for instance the presence of retrospective patterns. The review first examined general issues applicable to all stocks by considering how these were pertinent to three case studies (Georges Bank yellowtail, white hake and Georges Bank cod). The lessons learned from these then informed discussion on the assessment model to use for each of the 19 stocks.

The meeting opened on Monday morning (see agenda, appendix 3) with an overview of the current assessment approaches and challenges facing the 19 groundfish stocks, with consideration of working papers (appendix 4) on potential factors responsible for retrospective patterns and the model implications of zeros in surveys in the afternoon. Tuesday morning was devoted to consideration of the ability of statistical catch-at-age (SCAA) models to address issues raised, particularly causes of retrospective patterns, in the Georges Bank yellowtail, white hake and Georges Bank cod assessments. Tuesday afternoon was devoted to consideration of working papers examining the presence and influence of partial recruitment assumptions on the assessments. Discussion on Wednesday further explored the issues in the assessments (retrospective patterns in Georges Bank yellowtail domed partial recruitment in Georges Bank cod and aging errors in white hake). Based on the discussions during the first three days, on Thursday and Friday morning, there was stock by stock consideration by the Panel of the most appropriate formulations to use for each assessment. Unfortunately, an update on work by the Northeast Fisheries Science Center (NEFSC) to develop ecosystem models which will inform the next meeting on biological reference points, which has been planned for Thursday, could not be presented because of time limitations, although the text of the presentation was made available to the Panel. The last part of the meeting on Friday was devoted to the Panel reviewing its conclusions and discussing report assignments. The GARM review Panel consisted of José De Oliveira, Cynthia Jones, Paul Medley, Stratis Gavaris, Jim

Ianelli and Yan Jiao. The first three reviewers were assigned to the review by the national Center of Independent Experts (see statement of work for these CIE reviewers in appendix 6) while the last three were invited by the NEFSC. All were invited based upon their extensive expertise and experience with the issues considered by the meeting. The list of meeting participants is provided in appendix 5.

The presentation highlight of each working paper and the ensuing discussion as recorded by assigned rapporteurs is provided in appendix 7. These were important reference material to the Panel in drafting its report.

It is important to comment on the effectiveness of the meeting to address its terms of reference. Other than Terms of Reference 5, the panel was presented with sufficient material to address the stated terms of reference. Some clarification is required for terms of reference 5. For each of the 19 groundfish stocks, terms of reference 5 called for

> 'Assessment model that will be used to determine stock status and productivity characteristics until the next "benchmark" assessment is conducted. Where possible, apply the models to data (probably through 2006), to obtain current and historical estimates of F and B and estimates of uncertainty'

From the outset, it was apparent that in-depth review of the 19 assessments would not be possible. The approach taken at the meeting was to provide guidance on what class of model was most appropriate for each stock given the characteristics of the data (GARM III, 2007) and issues facing each assessment. Rather than focus on the software to be used, the Panel's review focused on the formulation most appropriate to take account of observational errors and model assumptions. These formulation decisions were considered more important than the choice of software because most applications have the flexibility to accommodate the formulation variants. While the details of each assessment formulation (e.g. ages to use in plus group) were not discussed, there was considerable discussion and guidance provided at the meeting that would inform some of these detailed decisions. Overall, while time was not available for in-depth review of the details of each assessment, the guidance provided by the Panel was important at this stage of development of the analyses in establishing the class of model to be used in each assessment. From this perspective, the Panel considers that important aspects of terms of reference 5 were met. If the NEFSC requires an in-depth review of the assessment formulations, this will have to undertaken later in the process.

The Panel was concerned about the limited time it had during the meeting to develop its conclusions. The Panel had planned for a full day of consideration of its report on the Friday but the meeting's agenda required the morning for completion of the discussion started on Thursday. While this requirement was understandable and perhaps unavoidable, review panels need time during meetings to consult. This needs to be factored into future GARM agendas.

Throughout the report, reference is made to a number of technical terms. The glossary provided in appendix 8 helps to clarify these terms.

PANEL RESPONSE ON TERMS OF REFERENCE

ToR 1. Applicability of one or more of the following modeling approaches to assess stock status (Index methods, Production Models, Age- or Length-based Models)

The Panel considers that model features required for the stocks considered in this GARM-III review must capture the underlying dynamics of populations and the key uncertainties about these and the associated data; capturing these uncertainties may require consideration of different model formulations. Models should be adequately conditioned on data so as to form the basis for the provision of advice. Appropriate consideration of uncertainty should allow risk assessment to be undertaken. Reference point generation and the ability to perform projections under alternative management options are also key features of prospective models.

A range of model classes was considered:

1. Relative trend models: These are models that consider trends in relative abundance and fishing mortality (F), and include the relative fishing mortality / replacement ratio approach of NEFSC (called "AIM"). The lack of catchability (q) estimates for this class of models limits its utility to inform management on absolute levels of allowable catch (i.e. TAC). They can however inform management on relative changes to existing TACs. Also, reference points have to be generated based upon expert judgment and on an *ad hoc* basis. There may be merit in investigating and comparing a more formal model fitting approach e.g. Glazer (2008, see next bullet), relative to the more exploratory, non-parametric approach of AIM.

2. Production models: These are essentially replacement yield models of the form: $B_{y+1}=B_y+Y^{rep}-C_y$[1], where Y^{rep} can take a number of forms, including a constant ($Y^{rep}=a$; Glazer, 2008), a linear ($Y^{rep}=a+bB_y$) or a quadratic ($Y^{rep}=rB_y[1-B_y/K]$; implemented in the NEFSC software ASPIC) form. These models can be used when age data are not available but have encountered problems when there has been depletion but no recovery in the time series. Caution is needed when using constant Y^{rep} for stocks for which there is no historical reported catch but there is evidence (e.g. anecdotal) that this could have been high or when the stock size has changed dramatically during the time period. In this situation, expert judgment may be needed to provide management with reference points based on these models.

3. Age-based models: These span the range from statistical catch-at-age to VPA-type models, and are the preferred approach when reliable age and growth data are available.

4. Length-and-age-based models: These are applied when age data are too limited to use age-based models, yet sufficient growth data (e.g. length frequency compositions) are available. However, it is important to evaluate stationarity when a growth model based on data from a restricted period is used.

Where data are lacking and there are conflicting hypotheses of underlying processes, the Panel noted the need to bound alternative states of nature by considering opposing extremes of plausible

---

[1] B_y = biomass in year *y*, C_y = catch in year *y* and Y^{rep} is the function describing the replacement yield

scenarios. Such an approach may require the consideration of multiple model formulations. If it leads to the consideration of competing models, then a risk analysis may be needed to inform decisions.

The software environment has important implications for what can be achieved. For example, a single software environment that can explore the diversity of data and assumptions is useful for comparative purposes. The use of simple models requires limited support, but also addresses a limited number of issues. On the other hand, more complex models require extensive, well-supported software (e.g. ASAP and SS2) that may limit broad access. Moderately complex approaches (e.g. ADAPT) require some support but are more widely accessible.

When moving from one class of models to another, or even from one formulation to another within the same class, it is important to document changes for fisheries managers. The Panel considers that this should be done by presenting results for the former approach (with new data), along with those for the new approach (with the same data). This allows the source of the change (whether new data or new model) to be clearly identified. It is important to provide rationale for any change in assessment models and explain differences in the results of the two approaches.

ToR 2. For certain stocks that are aged, utility of statistical catch-at-age vs. VPA based models with respect to Retrospective patterns, Flexibility to account for alternative parameterizations, Ability to incorporate external sources of information, especially tagging and environmental data and Ability to estimate parameters incorporating prior, external information

Age structured fish population models use observations on catch-at-age to determine population abundance-at-age. It is generally necessary to incorporate ancillary information to make this determination. Most commonly, the ancillary information comprises observations on abundance indices-at-age, e.g., from surveys or fishery catch per unit effort (CPUE). The catch and index observations may be used as amount-at-age or as split components of totals and proportions-at-age. The choice will influence how the error structures are handled. Assuming natural mortality is specified, a fully and freely specified age structured fish population model requires one population abundance parameter for each year class and one fishing mortality parameter for each age and year. It is generally not possible to reliably estimate all these parameters.

Two common approaches used to reduce the number of estimated parameters are referred to as Virtual Population Analysis (VPA) and Statistical Catch-at-Age (SCAA). VPA makes the assumption that the errors in the catch-at-age are negligible relative to the errors in other observations. This assumption may be held in situations where the catch is well monitored. Though not obligatory, to further reduce the number of parameters to be estimated, VPA frequently makes the assumption that the fishing mortality on the oldest age is a specified function, e.g. average, of the fishing mortality on younger ages. SCAA, on the other hand, makes the assumption that the fishing mortality is the product of an annual fishing mortality and an age specific partial recruitment (PR). This assumption may be approximated in situations where the nature of the fishery has not changed substantially over years. Also, the SCAA dissociates the catch and index data from the model whereas in VPA, the “catch data” are bound to the algorithm for computing numbers at age. Both VPA and SCAA are approximations to the complex reality and are suitable as a basis for providing management advice.

For partial recruitment structure, index relationship to population, handling of natural mortality, and ability to incorporate ‘prior’ information on estimated parameters (Bayesian techniques), both VPA and SCAA are equally flexible. The principal consideration when comparing VPA and SCAA is the tradeoff between the magnitude of the error in the catch-at-age and the stability of the partial recruitment pattern over years. When the error in the catch-at-age is considered substantial, SCAA may be preferred. When the partial recruitment pattern is likely to be variable from year to year, VPA may be preferred.

Often however, the results from VPA and SCAA do not differ substantially, suggesting that neither error in the catch-at-age nor departures from stable partial recruitment are critically important.

A more sophisticated variant of SCAA, referred to as semi-separable, relaxes the requirement for a stable partial recruitment pattern by allowing it to change over time. At the expense of having (usually many) more parameters, the semi-separable approach allows investigation to span the range from a VPA to a SCAA structure.

There are instances where, for some years of the analysis, the total catch is observed but the catch-at-age is unavailable. In these circumstances, only SCAA can be used to combine separate periods in the catch-at-age data. This may be of particular value for the estimation of reference points if there is an extensive initial period with total catch is observed but no catch-at-age data. Difficulties with a VPA application may also occur when errors in the catch-at-age are linked to some external or environmental observations.

An important issue in many fish stock assessments is the occurrence of retrospective patterns. These arise where the effect of removing recent data and re-running the assessment using earlier terminal years results in substantively different estimates compared to the result over the entire time period. These generally reflect incorrect model specifications or treatment of data and should be resolved. The panel reviewed results from a simulation experiment that investigated the performance of VPA and SCAA when retrospective agents were introduced and not accounted for in the model. The simulation operating model generated replicate observations which may have favored the SCAA approach because it used an underlying separable process for fishing mortality. A caveat on this is that the selectivity was length-based so the operating model was not entirely separable from an age perspective. Neither VPA nor SCAA were robust to the retrospective agents and their results displayed similar retrospective patterns. To properly address the retrospective patterns, it is necessary to know the true source, timing and magnitude of the effect (see section below on terms of reference 4). Uncertainty in point estimates should be carried forward to assist in judging the significance of the retrospective patterns. This may reveal that the estimates of uncertainty for one modeling approach contained the retrospective pattern residuals within confidence bounds more consistently than an alternate approach.

For many assessments reviewed at the meeting, the difference between results from a VPA and a SCAA approach were not the principal concern. Options for other model features had greater impact on results and deserve priority attention (see section below on terms of reference 5). The following table organizes these features and provides some guidance on the considerations for evaluation.

Model Feature	**Variants** (not comprehensive, reflects options considered at GARM III)	**Considerations**
Structural Assumptions		
Population Model	- annual time step (no within year dynamics) - homogeneous unit (no spatial sub-structure) - sex aggregated - sex disaggregated	- seasonal/spatial distribution of catches - sexual dimorphism
Natural Mortality	- assumed known and constant for all ages and years - age specific natural mortality	- demographics
Fishing Mortality	- separable as annual F and age specific PR	- reliability of total catch

	- separable as annual F and age specific PR, time block breaks/random walk	- accuracy and precision of catch age composition
	- separable as annual F and PR as function of age, time block breaks/random walk	- plausibility of constant PR for years in time block
	- separable as annual F and PR as function of length - calculated using catch equation, oldest age group PR estimated - calculated using catch equation, oldest age group PR conditioned on younger ages	- support for PR pattern/assumptions
Index Catchability	- proportional, time invariant within block of years - proportional, time block breaks	- support for change if applicable
Index Selectivity	- age specific, time invariant within block of years - function of age, time invariant within block of years - function of length, time invariant within block of years	- support for selectivity pattern
<u>Error Assumptions</u>		
Catch-at-age, expressed as catch numbers at age	- errors assumed negligible - Errors iid after log transform - Errors lognormal - Errors Poisson	- residual diagnostics - influential observations - magnitude of mean square residuals
Catch-at-age, expressed as total numbers and proportions at age	- errors assumed negligible on combined (total and proportion) - errors *iid* after log transform on combined - errors lognormal on proportion, assumed negligible on total - errors lognormal on proportion, lognormal on total - errors multinomial on proportion, assumed negligible on total - errors multinomial on proportion, lognormal on total - errors Poisson on combined	- residual diagnostics - influential observations -magnitude of mean square residuals
Indices at Age, <u>expressed as abundance at age</u>	- errors *iid* after log transform - errors lognormal - errors Poisson	- residual diagnostics - influential observations -magnitude of mean

		square residuals
Indices at Age, expressed as total and proportions at age	- errors multinomial on proportion, lognormal on total	
Penalty/Constraint	- none	- influence of constraints on estimates
	- recruitment conforms to Beverton-Holt form, log error, steepness<0.98 - recruitment conforms to a generalized Ricker form, log error - first year abundance conforms to equilibrium condition - on curvature of PR/catchability function - von-Bertalanffy growth	
Risk Analyses	- analytical confidence distributions	-confidence in measures of uncertainty
	- bootstrap confidence distributions - Bayesian posterior distributions - alternative states of nature	

ToR 3. Implications of zeros in the evaluation of fishery independent indices

When abundance at age estimates in NMFS surveys result in stratified mean values of zero, standard operating procedures are to treat these values as missing. This was the approach reviewed by the ICES WGMG 2007 report (ICES CM 2007/RMC:04). However in a NMFS Office of Science & Technology review in 2006, reviewers recommended an interim approach to handling these zeros as small values instead, pending further study. This problem arises when surveys used for tuning do not capture specific age classes. The presence of zeros is particularly problematic as age-stratified survey mean abundances are log-transformed before being used in VPA. Three working papers showed the results of using alternate approaches instead of changing zeros to missing. The papers explored several scenarios, including: 1) using actual values, or truncating values below a cut-off to zero and 2) treating zeros as missing, or replacing the zeros 3) an arbitrary small value, or 4) 1/6 times the smallest non-zero value. The simulations in working paper 3.1 showed that replacing zero with small values such as in scenario 3 or 4 was inappropriate because it resulted in bias. Prior to these simulations, previous studies (e.g. using g-statistic; Berry 1987) had identified a constant to replace zero based on obtaining normal skewness and kurtosis. However, any small constant is given considerable weight with a log-transform, whereas the Panel agreed that the zeros contain little information on stock size beyond abundance below some threshold. The Panel was in consensus that replacing zeros with arbitrary constants was not supportable because simulation results show that output can be biased when this is done. There was discussion as to how else to handle zeros as this poses a problem for models when abundances are low.

As an alternative to replacing zeros with an ad hoc value, the Panel discussed the potential of using other transformations beyond the log transform as a way to avoid problems with zeros. Several potential transformations of the Box-Cox type were briefly discussed, specifically the square root transform[2]. Such transformations can be used to stabilize the variance but may not always be appropriate; log transforms are from the same class of these transforms and have often been found to be suitable.

An additional approach would be to use other likelihood or quasi-likelihoods such as the Poisson or other distributions to handle zero-inflated models. Other distributions are used in modeling rare or elusive species and the Poisson has been previously used in fisheries. Multinomial (e.g. as implemented in the ASAP software), Conditional and Delta distributions have been used in other contexts and may or may not be appropriate for use when fitting populations trends from survey indices, depending on the situation. Changing to another distribution, such as the Poisson, has implications for q and variance estimation that would require further analysis and simulation. While some distributions, like the Poisson, can be fitted using iterative least squares, others, such as the negative binomial, do not have a simple relationship between the mean and variance and an additional dispersion parameter needs to be fitted as well as the mean.

The aggregation of age groups was seen to be advantageous particularly for older ages in which zero observations are more probable. One suggestion was to aggregate differently for any given year to better handle zero observations in these age groups. This has implications for the estimation of catchabilities at age (q) which would only be possible for consistently defined age groups over time.

The pattern of zero values is qualitatively useful in informing the assessment scientist that there may be a decline in the population or that there are issues with migration or some type of refuge, particularly as an adjunct to assessing the validity of using dome-shaped partial recruitment functions in the models. Even when data are omitted from the model fit due to the frequency of zeros, this pattern should be acknowledged and the original data presented.

The current NEFSC software which converts VPA to ASAP2 assumes a log normal likelihood for age composition data rather than a multinomial. The indices can also be entered in ASAP using the multinomial option, which could facilitate inter-model comparisons.

Altogether, the Panel considered that omitting zeros was a reasonable approach and that this was not a major issue for most assessments at this time. However, this issue could require further attention in the case of a declining sequence of biomass from surveys and would probably require identification of an alternative likelihood to the log-normal for survey data. As stated above, before being adopted, simulation should be conducted to explore the behavior of alternative formulations to address the presence of zeros in these data.

ToR 4. Potential factors responsible for retrospective patterns

Retrospective patterns result from structural errors in the stock assessment model. Where the error is consistent through time (e.g. a misspecification of natural mortality), a retrospective pattern will not occur. It will only occur when there has been a change within the time series of observations. There are four potential causes of retrospective patterns in age structured stock assessments:

- An unrecorded change in catches
- A change in natural mortality
- A change in the abundance index catchability (q)
- A change in fishery selectivity

---

[2] For example, a least squares general form would be $\sum\left(\sqrt{O_i}-\sqrt{E_i}\right)^2$ preserving the linear relationship between the observed and expected variables

In all cases, either the biomass has changed (changes in natural mortality and unrecorded catch) or is perceived to have changed (changes in catchability or selectivity) in a way that cannot be explained by the catch-at-age data, and therefore is a structural error in the model. Random noise is an unlikely cause, based on simulation analyses although it was hypothesized by a Panelist that mis-specification of the likelihood function could bring about retrospective patterns through influential data points.

One mechanism identified as a possible cause for retrospective patterns is when a survey of sessile species includes an area closed to fishing. When there is effectively no movement of the population between the closed and open areas, the part of the population in the closed area becomes unavailable to fishing and relatively more abundant than the part of the population outside the closed area. This causes an overall apparent change in the survey catchabilities. However, this is unlikely to be an important factor for groundfish.

The Panel notes that it is not possible to identify the cause(s) of the retrospective pattern from model diagnostics alone. Adjusting the model assumptions (e.g. altering survey q, catches or M) to remove the pattern does not guarantee the problem has been dealt with; the model may continue to be mis-specified. It is necessary to develop testable hypotheses concerning the cause, timing and magnitude of the effect. Additional information and analyses that might assist in discriminating between hypotheses include:

- Repeating the catch-at-age analysis on a “moving window” of time series data over the full time series to identify the timing of the cause;
- Examining the model residuals for a non-stationary pattern;
- Comparing trends in commercial and survey CPUE;
- Comparing the magnitude of survey catchability estimates across species and areas;
- Comparing swept area biomass estimates from surveys with those from the assessment model.

Some solutions to retrospective patterns were considered by the Panel, which in certain circumstances, the following could resolve:

- Truncate the assessment time series to the most recent period to remove the retrospective pattern. Although the underlying problem may continue to exist, it has been dealt with in a consistent enough manner for the time series in question to inform the decision–making. The consequences of doing this on reference point determination will require attention at the next GARM meeting.
- Consider the use of robust likelihood approach, which could address the hypothesized cases where influential data points are causing the retrospective patterns (Fournier et al., 1991).

In considering the recommendations of the NEFSC report of the retrospective working group, the Panel concluded that:

- A retrospective analysis should always accompany a catch-at-age stock assessment to determine whether or not a retrospective pattern is present.
- Hypotheses that might explain the cause(s) for the retrospective pattern should be developed and form the basis for further research; these need to be communicated to decision – makers as major sources of uncertainty.
- Standardized criteria need to be developed to identify a statistically significant retrospective pattern. The statistical test presented at the meeting, based on identifying a pattern that was unlikely to be caused by random error, looked promising. The properties and power of the test should undergo further examination.

- Where there are strong grounds to reject the best single model, the alternative models (hypotheses) which explain the retrospective patterns should be used to evaluate the consequences of management decisions. The uncertainty needs to be communicated to managers and decision-makers, through decision tables and other types of risk analysis.

ToR 5. For each stock, assessment model that will be used to determine stock status and productivity characteristics until the next "benchmark" assessment is conducted. Where possible, apply the models to data (probably through 2006), to obtain current and historical estimates of F and B and estimates of uncertainty

General Considerations

In this section, for each stock, the Panel provides guidance on the class of assessment model that are suitable given the nature and availability of the relevant data. Often, there was no compelling evidence to choose between a VPA and SCAA formulation. The analyses presented at the meeting using comparable VPA and SCAA formulations generally behaved similarly. In situations where the Panel considered that one formulation was more suitable, the preference is stated.

Following this, the most influential model features that need to be resolved for each stock are stated in order of priority. These are generally, but not always, intended to address retrospective patterns observed in current assessment formulations.

A number of issues arose that are relevant to all stocks and are discussed first.

Estimation of Partial Recruitment

The estimation of the partial recruitment pattern on the older age groups was a recurring issue in a number of the stocks. However, while dome-shaped fishery partial recruitments on the older age groups may resolve retrospective patterns, it may also lead to what was termed 'cryptic' biomass – biomass generated by the model that has not been observed in either the fishery or surveys.

The Panel considers that the burden of proof should be to convincingly demonstrate that the fish exist in the population when not observed in the fishery and surveys, even if the model fit with dome-shaped partial recruitment is better. Patterson (2002) noted for highly parameterized models, fishing mortality tended towards zero while population numbers tended towards infinity. Other solutions can only be achieved by imposing some constraints on the models e.g. constraining fishing mortality on the oldest age group in each year to a fixed proportion of an average of fishing mortality over younger ages as in ADAPT and Laurec−Shepherd methods. This is another way of saying that additional information (data and / or assumptions) external to the model is required. At the very least, the consequences of adopting a dome-shaped as opposed to a flat topped partial recruitment should be documented for consideration in the management arena. When competing hypotheses of partial recruitment are being considered, all information, including external sources of data, should be examined to inform the merits of each. For example, it may be possible to disaggregate the plus group catch to assist in hypothesis testing. Analyses of tagged fish presented at the meeting offered promise for evaluating hypotheses about partial recruitment. A number of different hypotheses emerged at the meeting that could be formally investigated which the Panel encouraged.

When considering an SCAA formulation, it is often necessary to define blocks of time during which the separable assumption is likely to be met. However, care needs to be taken on how to define the transition between blocks. Abrupt changes in estimated partial recruitment may result from large errors in the observed data. Techniques (e.g. random walks) which constrain how much partial recruitment can vary between adjacent blocks of time should be explored.

The algorithm in the software of the NEFSC used to estimate the population numbers in the plus group gives rise to inconsistencies. It assumes that the fishing mortality on the plus group is some function of the fully-recruited fishing mortality in the same year and, using the Baranov catch equation, estimates plus group abundance from the catch numbers of the plus group. This does not recognize abundance contributions of non-plus group and plus group ages from the year previous. An appropriate algorithm is provided in Anon. (2003). Differences between the two algorithms depend upon the size of the plus group as well as the assumptions made on the partial recruitment.

While the difference between the two algorithms is small when plus group catch is small, it can be important when plus group catch is large and / or when the partial recruitment is assumed to be domed.

Weighting of Model Components

When fitting models to data, weighting factors are often used to emphasize or de-emphasize components, e.g. less weight given to the index for age 1 relative to others because it is considered 'noisy'. The Panel did not extensively explore how to define weights but offers some observations, based upon discussion at the meeting.

Weighting may be internal or external. In the first case, the weights applied to the fit of each model component are based on the measure of fit of the model to the data and are iteratively updated during the estimation process. In these situations, it may be useful to define a minimum bound on the weighting (e.g. minimum variance) to ensure that no one model component draws an inordinate amount of the weighting.

External weightings may be derived from estimates of the variance in the input datasets (e.g. stratified variance in the survey dataset). In these cases, it is important to ensure that the estimate of the variance is precise enough to be useful. Small sample sizes can lead to imprecise estimates of the variance. It may be possible to generate externally derived weights based on a statistical analysis of trends in variance in the dataset (i.e. produce smoothed estimates of the variance used for weighting). This would have to be justified on a case-by-case basis. Another consideration is when the sampling variance does not fully characterize the total error, which may result in inappropriate weighting of components. This situation could be rectified by introducing an additional variance parameter to be estimated (Germont and Butterworth, 2001; Punt et al., 1997).

Ad hoc external weighting (choice of weights based upon non-statistical arguments) should be avoided. If expert judgment is required to establish external weightings, the rationale should be clearly documented and the impact of the choice should be described. Experimentation with alternative weighting may be a useful tool for exploration of the influence of components.

During the meeting, two different ways to implement external weighting were discussed. The first involved multiplying the component's measure of model fit (least squares or likelihood) by a constant (lambda) while the second involved dividing the fit by a constant, typically a measure of the variance of the dataset under consideration. In some cases, both types of weighting were employed. The Panel considered that either approach, but not both simultaneously, is appropriate. Use of both approaches in the same model fitting process has the potential to confuse determination of what the real weighting is. It was noted that many assessment packages (e.g. ASAP, SS2) use the lambda constant to turn on (=1) or off (=0) specific components of the objective function, which is an appropriate use of the term.

Biological Reference Points (BRPs)

The relative merits of estimating biological reference points internally (within the model fitting process) or externally (analyzing the assessment results to produce the BRPs) were discussed. One of the main considerations in determination of BRPs is characterization of the stock - recruitment relationship. Often the variability about this relationship is very high. Accordingly, the weighting of this component in the model fitting is low so that determining its parameters internally will not greatly influence estimation of current stock size. Its real value is in estimation of BRPs. If the same data are used, internally (assuming non-informative priors) and externally derived BRPs would be expected to be similar.

A potential advantage of estimating BRPs internally, say within an SCAA formulation, is that it allows inclusion of years in which only catch data are available. This allows consideration of years of data early in the history of exploitation of a stock which can provide a perspective of long-term productivity. This potential advantage has to be weighed against the possibility that productivity has not remained stationary over such extensive time periods.

Estimating BRPs externally is a useful way to corroborate the BPRs estimated internally using the same dataset and is encouraged by the Panel. If differences between the two approaches are encountered, these should be investigated and explained.

Stock-by-Stock Assessment Formulations

The ordering of the stocks below conveys the Panel's overall sense of priority to address assessment issues. This is not the priority for undertaking the assessment which may be driven by management need. Rather, it is based on the Panel's understanding of the status of the current assessment formulation and the likelihood of improvements being made to it.

Gulf of Maine Cod

Fishery sampling and aging data are generally good. While the recreational catch contributes larger errors, on balance, the data are sufficient to employ an age – structured model assuming negligible errors in the catch-at-age.

There is only a weak retrospective pattern using the current VPA. However, there is a need to confirm the partial recruitment on ages five plus as this is particularly influential on the estimation of biological reference points and on stock status determination. The analyses presented at the meeting (working papers 2.2 and 2.3) indicated that the model fit to the data favored a domed partial recruitment. Resolution of the differences in the results with alternative partial recruitment patterns call for hypotheses on causative processes as well as external sources of information. Regarding the latter, the shape of the partial recruitment curve for the older ages could be investigated through more detailed examination of the dynamics within the plus group (age 7+) as well as analysis of the available tagging data. The consequences of either hypothesis on management advice should be explored through a risk analysis.

Southern New England, Mid-Atlantic Yellowtail Flounder

Given the good sampling rates and availability of aging information, the data are sufficient for an age-structured model. The Panel felt that negligible error in the catch-at-age could be assumed although this assumption may need to be relaxed for the recent time period.

There is a moderate retrospective pattern in the current VPA formulation that requires examination. Following the approach for the Georges Bank stock, it would be worthwhile splitting the survey time series to explore whether or not similar trends in survey catchability are present in Southern New England. If so, splitting of the survey time series could be a proxy for changes in survey catchability related to habitat use by yellowtail. Yellowtail might be occupying more preferred habitat due to environmental or other influences, causing survey catchability to increase. This hypothesis should be investigated.

The Panel noted the need to investigate long-term changes in stock productivity given the severe decline of the resource. This has implications for the determination of biological reference points. It could not comment on the details of this examination.

Cape Cod-Gulf of Maine Yellowtail Flounder

The Panel comments made for Southern New England / Mid-Atlantic yellowtail are also applicable to the Cape Cod, Gulf of Maine yellowtail stock.

Witch Flounder

Coefficients of variation on the landings-at-age overall are about 17%, there are no recreational catches and discards are low. The data are sufficient to use an age-structured model assuming negligible error in the catch-at-age. However, comparison with an SCAA model that allows for error in the catch-at-age may be useful in this case, following resolution of the retrospective issue.

While the current VPA formulation has been adequate in capturing the broad-scale dynamics of the stock, it has exhibited a consistent retrospective pattern that warrants exploration. The Panel could not comment on the nature of these explorations other than point out the potential sources of retrospective patterns made elsewhere in this report and the need to bring external sources of data to bear on identifying potential causative processes.

The Panel noted the desire of the NEFSC to use the ASAP software package to undertake these explorations. There are no compelling reasons not to switch from ADAPT to ASAP if this makes analyses easier.

Redfish

The data are sufficient for application of an age-structured model. This is important given the strong evidence for infrequent large pulses of recruitment which persist in the stock over decadal time periods. The Panel could not evaluate if error in the catch-at-age could be assumed negligible. This should be examined to determine if a SCAA approach would be more suitable.

The fishery may target abundant year-classes as they move through the stock. It will be important in the assessment to relax the separable assumption to allow for this possibility and explore if it is occurring.

In relation to biological reference points, internal estimation of the stock - recruitment relationship in a SCAA formulation will need to take account of both the stock's inherent productivity and the presence of episodic large year-classes. In relation to the former, steepness (h) can be inferred from redfish resources elsewhere (e.g. West Coast). In relation to the latter, the analysis should consider setting the assumed error around the stock - recruitment relationship ($\sigma_{R,}$) at a low constant value (e.g 0.2 or less) for years when there is limited age data (i.e. little information on year-class strength) and then to increase it to an appropriate higher constant value (e.g. 0.4 or higher) for periods when age data are more plentiful (Maunder and Deriso, 2003). The Panel considers that the sensitivity of the assessment to these stock - recruitment assumptions should be checked through comparison of the model results without them. In addition, the Panel considered that an externally-derived surplus production model should be investigated to evaluate the robustness of derived biological reference points.

Georges Bank – Gulf of Maine White Hake

The differentiation between red and white hake in the commercial fishery is an issue for this stock, particularly for discards which in some years can represent 50% of the catch, much of this being less than 60cm in length. While there is no commercial aging data since 2000, length frequency sampling is available for this period. Thus the data are sufficient for an age-structured model although negligible error in the catch-at-age cannot be assumed. A model should be used that can take advantage of the age and length frequency data available (e.g. SCALE). There is a potential for sexual dimorphism to confound this attempt at modeling, which should be considered in the model formulation if possible.

The species identification problem in catch samples for lengths less than 60 cm can be examined by consideration of species composition for these fish sizes in the survey dataset, calibrating these with available observer data.

Georges Bank – Gulf of Maine American Plaice

The data appear to be sufficient to undertake an age-structured model. However, the Panel could not evaluate whether or not the error in the catch-at-age could be assumed to be negligible. Since the discards are an important fraction of the catch and appear not to be well-determined, it is appropriate to assume the presence of error in the catch-at-age (e.g. SCAA).

The current VPA formulation exhibits a moderate retrospective pattern, the causes of which require examination.

There is a potential problem of conducting an assessment on the combined Georges Bank and Gulf of Maine stock subcomponents if the relative proportion of abundance of these stocks is not stable over time. The survey trends in the two areas should be examined; if they are similar, then a combined assessment of the two components should not be problematic. However, if the trends are different, there may be a need to partition the catch-at-age between the two stocks and conduct separate assessments on each assuming that there is negligible migration between the two populations.

Plaice growth rates have been observed to be different on Georges Bank and in the Gulf of Maine. This has consequences for the partial recruitment in each area. There is a need to validate the separable assumption of the SCAA when undertaking the assessment.

Georges Bank Winter Flounder

The Panel noted the improvement in commercial age and length sampling since 2000 although gaps in these data exist in the middle of the time series (1998-99). The data are sufficient to undertake an age - structured model but error in the catch-at-age may not be negligible. Thus, the stock is a candidate for a SCAA formulation.

The Panel could not assess the overall utility of a SCAA formulation without results to examine. The Panel recommended that one model approach be chosen and if there are no problems observed then there is no need to explore an alternative approach.

The Panel noted similar issues with the other two winter flounder stocks (Gulf of Maine and Southern New England) While each has its challenges regarding data availability and other issues, the NEFSC should consider whether or not a common class of models could be applied to all three stocks. This will not only facilitate comparison of stock dynamics but also lead to new insights in each assessment.

Gulf of Maine Winter Flounder

It was noted that year-classes that can be identified in the commercial and survey length compositions do not appear in the respective age composition. This may indicate smearing across year-classes in the age-length keys. As well, commercial sampling intensity has been low. Thus, while the data appear sufficient for an age-structured model, negligible error in the catch-at-age cannot be assumed. In addition, a formulation that takes advantage of the available length frequency information would be appropriate (e.g. SCALE).

The current VPA formulation exhibits a strong retrospective pattern. There was a suggestion that the causes of the retrospective might be related to sexual dimorphism, which could be investigated through a sex-separate model, although the report of the NEFSC Retrospective Working Group (working paper 4.1) indicated that a strong retrospective pattern could not be generated by sexual dimorphism alone. It will also be useful to corroborate model output by comparing fishing effort trends implied by the models with reported fishing effort.

Conflicting patterns between the recruitment and biomass time series were noted. To represent these inconsistencies in the data, a risk evaluation should be undertaken, which compares results between models weighted towards either the adult or the recruitment indices. The consequences of each weighting on harvest advice should be documented.

Southern New England Winter Flounder

The data appear to be sufficient for an age-structured model and information available suggests that negligible error in the catch-at-age can be assumed.

The current VPA formulation exhibits a strong retrospective pattern but seems to be transitory as it is not evident in recent years. Some event may have occurred in the mid 1990s to cause the retrospective pattern that is no longer occurring.

The causes of the historical retrospective patterns should be explored using the current assessment formulation. A possible change in the catch-at-age around 1994 could be checked to see if the retrospective was due to a single event related to this. If the retrospective pattern does not reappear with the additional years of data to the assessment, then there is no need for further exploration. If on the other hand, it does reappear, consideration should be given to splitting the survey time series pre and post 1994. Further, as discussed in the section on terms of reference 4, a robust likelihood approach might address the retrospective pattern if some of the indices have been very influential.

Georges Bank – Gulf of Maine Pollock

While there are issues with determining the age structure of the commercial catch in the Canadian and distant water fleet components in the 1960s and early 1970s, the data may be sufficient for use of an age-structured model post 1977. The Panel could not evaluate if the error in catch-at-age was negligible.

The Panel considered that the Relative Trend class of models is likely informative given the strong relationship between the relative fishing mortality and replacement yield for this resource and thus could be the basis of the 2008 assessment.

The Panel noted that suspected transboundary US / Canada migration will impact the assessment. Joint research on stock structure with the Canadian Department of Fisheries and Oceans has been proposed (TRAC, 2007).

The aging data may be informative about recruitment and could augment the Relative Trend analyses. This information should be evaluated and presented at the April GARM III review.

Georges Bank – Gulf of Maine and Southern New England- Mid-Atlantic Windowpane Flounder

Except for the fall 1999 NMFS survey, there is no commercial aging information for these stocks. Commercial and survey length frequency data are available that allow examination of the dynamics of recruitment and a plus group separately within an age structured model. A high percentage of the catch has been discarded, particularly since 2000 when it has been in excess of reported landings. It may not be appropriate to assume negligible errors in partitioning the recruitment and plus group, in which case SCAA would be more suitable.

While the current Relative Trends approach (AIM) appears to be adequate for the SNE-MAB stock, the Panel recommended exploring the use of SCAA in a Collie-Sissenwine formulation as the assessment model for the GOM / GB stock. There are benefits to using a common assessment framework for both stocks. Therefore, future assessments should consider a common model formulation.

Georges Bank Yellowtail

The data for this stock are sufficient to undertake an age - structured model assuming negligible errors in the catch-at-age.

A previous VPA formulation had exhibited a strong retrospective pattern. The Transboundary Resource Assessment Committee (TRAC, 2005) examined the causes of the retrospective pattern in detail and while not identifying the primary cause, developed an assessment formulation (termed the Major

Change Model) that, through splitting the survey time series pre and post 1994 largely removed the pattern and was considered suitable as the basis for management.

There was nothing presented at this GARM III meeting which would lead to improvement of the Major Change Model and thus it should be the basis of the GARM III assessment.

Notwithstanding this, the primary cause of the retrospective pattern remains unresolved. As noted for SNE / MAB yellowtail, splitting of the survey time series could be a proxy for changes in survey catchability related to preferred habitat use by yellowtail, motivated by environmental or other influences. This hypothesis should be investigated for both stocks. If a mechanism for the apparent change in survey catchability in the mid- 1990s can be identified, this should be presented along with the current Georges Bank model.

While the partial recruitment pattern on the ages four-plus could be investigated with the Major Change Model, this issue was not identified as a concern during the TRAC benchmark review (TRAC, 2005). Model fits presented at this meeting that had not adopted the Major Change Model formulation, suggested dome partial recruitments in both the survey and commercial fishery. However, that analysis concluded that the problems indicated by the strong residual patterns, and most likely a strong retrospective pattern, should be resolved before considering the results on the partial recruitment. Further, a domed fishery partial recruitment is at odds with the results of tagging analysis presented at this meeting, which suggested no dome. There was discussion at the meeting on the potential confounding of the tagging results by other processes (e.g. gear avoidance, emigration, tag reporting rate), and although none of these appeared likely, there was insufficient time at the meeting to fully explore the alternatives. Thus, while a domed partial recruitment on the older age groups appears unlikely, further exploration of external data could be undertaken to corroborate this.

If a SCAA formulation is considered, conducting the analysis using the catch and discards by separate fleet components may improve model fits as the separable assumption is more likely to be met within these.

Finally, overall improvements in the stock assessment of yellowtail may be gained by considering all three stocks (Georges Bank, Cape Cod – Gulf of Maine and Southern New England – Mid Atlantic) as a complex with migration between components. This is not suggested for immediate exploration.

Georges Bank Cod

The data are sufficient for an age-structured model in which it can be assumed that the error in the catch-at-age is negligible.

The current VPA formulation exhibits a weak to moderate retrospective pattern, the presence of which may be due to changes in the fishery partial recruitment since 1994. This may be related to major changes in management structures (e.g. closed areas, mesh size) made at the time. Model formulations that allow investigation of changes in partial recruitment, particularly with regards to the older age classes in the recent time period (post 1995) should be pursued.

Georges Bank Haddock

The data are sufficient for an age-structured model which assumes negligible error in the catch-at-age. It is noted that catch-at-age data for this stock is available as a continuous time series as far back as 1930 (Clark et. al, 1982). These data should be included in the assessment.

The current VPA formulation exhibits only a weak retrospective pattern. However, recent changes in haddock size at age (declining) have implications for the assumption of stationarity of survey catchability at age and for the estimation of BRPs which will have to be addressed in any age structured model. If a SCAA formulation is considered, a number of partial recruitment blocks throughout the assessment's time period will likely be required.

Regarding biological reference points, in addition to the general comments made above, the last time that a stock – recruitment relationship was examined for this stock (NEFSC, 2002), a non-parametric form had to be used due to a lack of convexity in the relationship. This will likely persist, suggesting that the relationship might better be considered externally to the assessment model.

Gulf of Maine Haddock

The data may be sufficient to undertake an age - structured model although it has not yet been processed. The Panel encourages completion of the processing of the relevant data as time is available. If this information is not processed before the next assessment is due, a Relative Trend class of models will have to be employed (e.g. AIM).

The spatial distribution of the catch and the surveys should be examined to determine whether or not the high landings and survey observations in the Great South Channel area are spill over from the Georges Bank stock. This may affect the perception of the productivity of this stock.

Ocean Pout

Recent landings are bycatch in other fisheries. While length frequency sampling for discards has been good in recent years, there are no aging data available. Consequently, the data are not sufficient for an age-structured model.

While the current Relative Trends approach (AIM) could be used for the GARM III assessments, alternative models should be explored that both have a stronger basis in biology and more explicitly address uncertainty. These include age-structured models using life history parameters derived from literature and other external sources as well as Bayesian biomass dynamics models. The sufficiency of available age data to construct the growth relationship to support these models requires exploration.

There is only a weak relationship between Relative Fishing Mortality and Replacement Ratio, suggesting that the Relative Trends class of assessment models is not informative for reference points. It is possible that the stock's dynamics have been severely impacted by fishing historically, to the point that it may not be possible to determine the link between exploitation rate and productivity.

Atlantic Halibut

There is a long time series of landings data for this stock that could inform estimation of its productivity. Sampling intensity has been low and while length frequencies have been collected, there are no aging data. The data are not sufficient for an age-structured model.

The Panel suggests attempting a one parameter (for productivity) depletion analysis assuming a plausible landing level before 1893 and fit to available survey abundance trends. Notwithstanding this, 28 percent of halibut tagged in US waters are returned from the Canadian zone suggest that this stock should be assessed in collaboration with Canadian assessment scientists.

ToR 6. Sufficiency of the assessment models to estimate measures of stock status consistent with Biological Reference Points

There are model approach implications for methods to estimate biological reference points. The Panel spent little time evaluating assessment model considerations relative to BRPs directly, although a number of issues related to this were raised throughout the meeting. Some of the key issues identified were:

- Length of time series considered (the problem of shifting baselines)

- Assessment of stock status generally relates to the most recent time period while the estimation of BRPs requires a longer term perspective of productivity
    - Inclusion of historical catch data has implications for BRPs and long-term changes in stock structure needs to be considered
    - Temporal changes in biological parameters (e.g. recent declines in growth) need to be taken into account
    - For long catch time series (e.g., redfish) deriving an appropriate treatment for historical recruitment pattern consistent with removals is critical
- The estimation of stock - recruitment relationships internally and externally to the assessment models (see also comments on BPRs in section on terms of reference 5)
    - VPA and SCAA can estimate population stock-recruitment relationships integrated with survey and catch data and have the potential to estimate BRPs; SCAA can extend this analysis to years in the absence of a complete time series of catch-at-age data
    - It is desirable that the approach selected provides reasonably consistent estimates of BPRs (i.e., inter-assessment variability should be low)
    - Stock-recruitment assumptions are key for the estimation of biological reference points but less critical to the determination of current stock status. Also, the flexibility provided by both parametric and non-parametric estimation approaches should be considered
    - Where possible, the Panel encourages estimating BRPs both internally and externally, comparing the results and explaining differences
- The potential for ambiguity when BRPs are estimated using expert judgment (e.g. biomass reference points for stocks assessed using AIM). The Panel recommends the development of a consistent and defensible basis for estimating BRPs using the Relative Trends class of assessment model. This includes a transparent basis for estimating scale
- The extent that a stock can be considered adequately managed as a "unit" (i.e., the area is at the margins of the natural range for a stock - e.g., Atlantic halibut)
- Overall multispecies production (to be more fully addressed at the BRP GARM).
    - Technical interactions (bycatch of co-occurring species) methods need to be pursued for practical management applications

The above points need to be kept in mind when applying assessment methods for estimating BRPs.

The Panel suggested an approach for age-structured models to evaluate the impact of fishing on a stock under minimal assumptions about the underlying productivity (e.g., stock-recruitment relationships). Specifically, the historical time series of recruitment estimates from a particular assessment could be used to compute the subsequent spawning biomass levels as if no fishing had occurred. If a stock-recruitment relationship is estimated, the original recruitment estimates can be adjusted by the ratio of the expected recruitment given spawning biomass (with and without fishing) and the estimated stock-recruitment curve i.e. the recruitment under no fishing modified as:

$$R_t^{'} = \hat{R}_t \frac{f\left(S_t^{'}\right)}{f\left(\widehat{S}_t\right)}$$

where $\hat{R}_t$ is the original recruitment estimate in year t with $f\left(S_t^{'}\right)$ and $f\left(\widehat{S}_t\right)$ representing the stock-recruitment function given spawning biomass under no fishing and under the fishing scenario, respectively. This approach would be particularly useful where the stock-recruitment relationship is of the Ricker type, where recruitment begins to decline at higher stock biomass.

For example, application to Gulf of Maine cod using reference case estimates of recruitment, maturity and mean weights-at-age from the meeting working papers, are projected forward from age one but *without F*. Without any stock-recruitment effect, a current level of "depletion" is about 26% (figure 1). Adding an adjustment due to a stock-recruitment relationship is likely to change this pattern to some degree, particularly if a Ricker stock – recruit model is used, as noted above. This technique can be used to test both the impact of fishing and the assumptions about the stock - recruitment relationship.

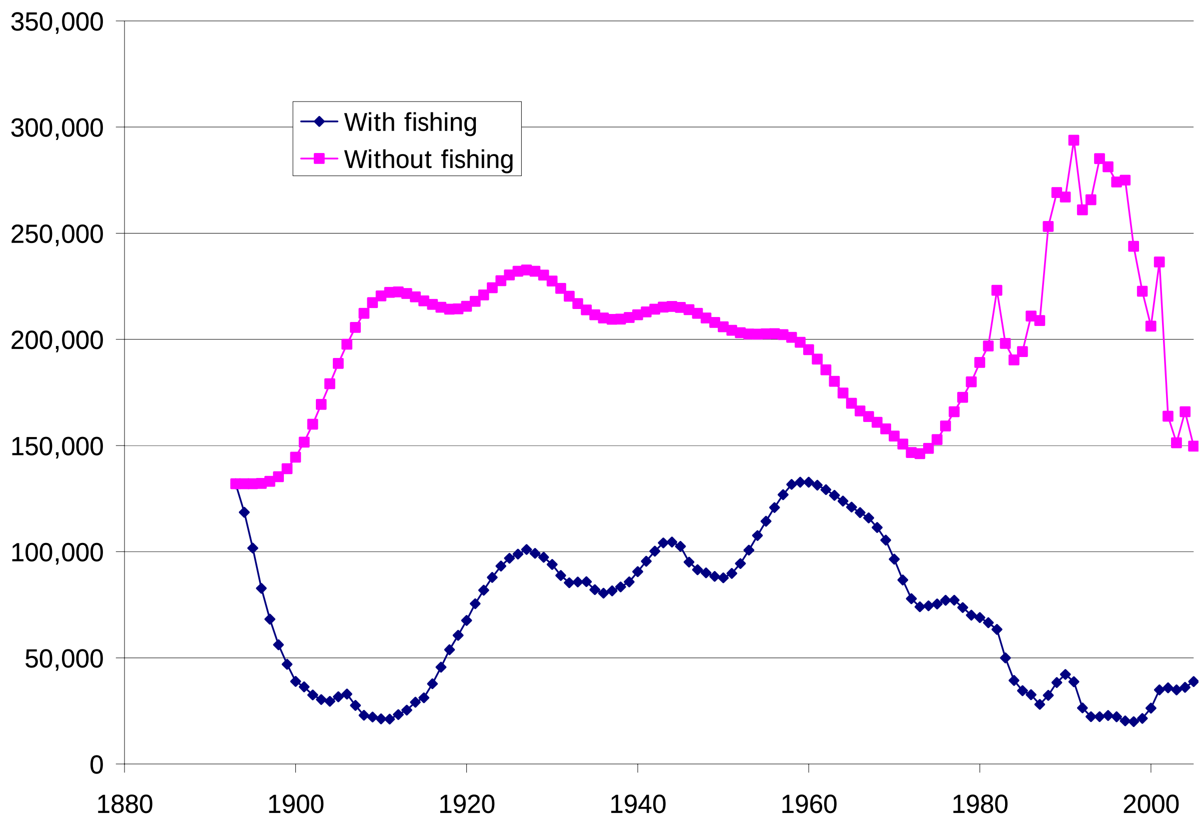

Figure 1. Historical spawning stock biomass for Gulf of Maine cod based on the reference model (with fishing) compared to estimates assuming natural mortality of 0.2 only (without fishing)

CONCLUDING REMARKS

The meeting required an extensive suite of working papers prepared by scientists at the NEFSC and substantial and in-depth discussions at the meeting itself. This was a very significant workload by the Center, which the Panel acknowledges being of very high quality. The Panel would also like to acknowledge the valuable contributions at the meeting made by all participants, particularly those of Doug Butterworth and Rebecca Rademeyer, who attended on behalf of the fishing industry. Finally, the Panel would like to thank Andrea Strout of the NEFSC who assisted the chair in preparing this report. All these contributions made it possible for the GARM III 'models' review to meet its terms of reference.

REFERENCES

Ames EP. 2004. Atlantic Cod Stock Structure in the Gulf of Maine. Fisheries. 29: 10 – 28.

Anon. 2003. Workshop on Assessment Methods. NAFO Scientific Council Studies. 36: 215 – 216

Berry DA. 1987. Logarithmic transformations in ANOVA. Biometrics 43:439-456.

Brandao A, Butterworth DS. 2008. An updated assessment of the South African Kingklip resource including some sensitivity tests. Marine and Coastal Management document. WG-Dem: K:02 revision2

Butterworth DS, Ianelli JN, Hilborn R. 2003. A Statistical Model for Stock Assessment of Southern Bluefin Tuna with Temporal Changes in Selectivity. Afr. J. Mar. Sci. 25: 331–361.

Clark SH, Overholtz WJ, Hennemuth RC. 1982. Review and assessment of the Georges Bank and Gulf of Maine haddock fishery. J. Northw. Atl. Fish. Sci. 3: 1 – 27.

Fournier DA, Sibert JR, Terceiro M. 1991. Analysis of length frequency samples with relative abundance data for the gulf of Maine northern shrimp,(*Pandalus borealis*) by the MULTIFAN method. Can. J. Fish. Aquat. Sci. 48: 591-598.

Germont HF, Butterworth DS. 2001. Possible extensions to the ADAPT VPA model applied to Western North Atlantic Bluefin Tuna, addressing in particular the need to account for "additional variance". ICCAT Coll. Vol. Sci. Pap. 52: 1663-1705.

Glazer J. 2008. Updated replacement yield estimates for the South African Monkfish resource, *Lophius vomerinus,* and analyses of trends in abundance. Marine and Coastal Management document. WG-Dem:A:01 revision1.

Herrington WC. 1935. Modifications in gear to curtail the destruction of undersized fish in otter trawling. US Bureau of Fisheries Report. 46 pp.

Patterson, KR. 2002. Exploring and quantifying structural uncertainty in age−structured fish stock assessment: an approach based on a "kernel" survivors analysis. ICES C.M.2002: V 10.

Maunder MN, Deriso RB. 2003. Estimation of recruitment in catch-at-age models. Can. J. Fish. Aquat. Sci. 60: 1204–1216.

NEFSC 2002a. Assessment of 20 Groundfish Stocks through 2001. Northeast Fisheries Science Center Reference Document 02-16.

NEFSC. 2002b. Re-Evaluation of Biological Reference Points for New England Groundfish. Northeast Fisheries Science Center Reference Document 02-04.

NEFSC. 2005. Assessment of 19 Northeast Groundfish Stocks through 2004. Northeast Fisheries Science Center Reference Document 05-13

NEFSC. 2007. Report of the Groundfish Assessment Review Meeting (GARM III). Part 1. Data Methods. NEFSC Report. 69 pp.

PUNT AE, Cooke JG, Borchers DL, Strindberg S. 1997. Estimating the extent of additional variance for Southern Hemisphere minke whales from the results of the IWC/IDCR cruises. Rep. int. Whal. Com. 47: 431-434.

Serchuk FM, Wigeley SE. 1992. Assessment and Management of the Georges Bank Cod Fishery: An Historical Review and Evaluation. J. Northw. Atl. Fish. Sci., Vol. 13: 25–52.

TRAC. 2002. Proceedings of the Fifth Meeting of the Transboundary Resources Assessment Committee (TRAC), Woods Hole, Massachusetts, February 5-8, 2002. Northeast Fisheries Science Center Reference Document 02-12.

TRAC. 2005. Proceedings of the Transboundary Resources Assessment Committee (TRAC) Benchmark Review of Stock Assessment Models for the Georges Bank Yellowtail Flounder Stock. Transboundary Resource Assessment Committee. Proceedings 2005 / 01.

TRAC. 2007. Canada USA Scientific Discussions. 10 – 12 April 2007. Transboundary Resource Assessment Committee. Proceedings 2007 / 02.

APPENDICES

Appendix 1. Terms of Reference for the GARM-III Models Meeting

1. For each stock, consider the applicability of one or more of the following modeling approaches to assess stock status:
    - Index methods
    - Production Models
    - Age- or Length-based Models

2. For certain stocks that are aged, compare and contrast the utility of statistical catch-at-age vs. VPA based models with respect to the following criteria:
    - Retrospective patterns
    - Flexibility to account for alternative parameterizations
    - Ability to incorporate external sources of information, especially tagging and environmental data
    - Ability to estimate parameters incorporating prior, external information.

3. Address the implications of zeros in the evaluation of fishery independent indices.

4. Examine potential factors responsible for retrospective patterns.

5. For each stock, define the assessment model that will be used to determine stock status and productivity characteristics until the next "benchmark" assessment is conducted. Where possible, apply the models to data (probably through 2006), to obtain current and historical estimates of F and B and estimates of uncertainty.

6. Evaluate the sufficiency of the assessment models to estimate measures of stock status consistent with Biological Reference Points.

Appendix 2. Draft Terms of Reference for the GARM-III "Biological Reference Point (BRP)" Meeting (28 April – 2 May 2008)

1. For relevant stocks, determine the influence of retrospective patterns in parameter estimates (e.g., fishing mortality, biomass, and/or recruitment) from assessment models on the computation of BRPs and on specification of initial conditions for forecasting

2. Trends in Stock Productivity:

   a.) For relevant stocks, identify trends in biological parameters (i.e., life history and/or recruitment) and assess their importance for the computation of BRPs and for specification of rebuilding scenarios

   b.) If possible, summarize trends in pertinent environmental variables that might be related to the trends in those biological parameters relevant to BRPs

3. Ecosystem approaches to Gulf of Maine/Georges Bank fisheries:

   a.) Determine the production potential of the fishery based on food chain processes and estimate the aggregate yield from the ecosystem

   b.) Comment on aggregate single stock yield projections in relation to overall ecosystem production, identifying potential inconsistencies between the two approaches

4. Biological Reference Points (B_{target}, $B_{threshold}$, F_{target}, $F_{threshold}$):

   a.) For each stock, list what the <u>current</u> BRPs and/or BRP Proxies are (e.g., B_{MSY}, B_{MAX}, F_{MSY}, $F_{40\%MSP}$, historical survey catch per tow, etc.), and give their values (i.e., typically from GARM II)

   b.) For each stock, update or redefine BRPs or BRP proxies that will be used for stock status determination, and compute their expected values and precision. Note: These BRPs and their proxies must be comparable and consistent with outputs from the recommended assessment models from the GARM III "Modeling" Meeting

5. For each stock, identify appropriate models for forecasting and for evaluating rebuilding scenarios.

Appendix 3. Draft Meeting Agenda of GARM III Models Meeting. Feb. 25-29, 2008, Stephen Clark Conference Room, NEFSC, Woods Hole, MA

Monday February 25

0900-0910 Welcome (Deputy Director)
0905-0910 Introductions
0910-0940 Overview of GARM and objectives of this meeting (Chair)

TOR #1: Applicability of Models to Assess Stock Status

0940 -1020 Overview of GARM species, data availability, assessments
Working Paper 1.1, 1.3 - (Rago)
1020-1040 Discussion
1040-1055 Break
1055-1125 Review of modeling approaches and rationale
Working Paper 1.2 - (Rago)
1125-1200 Discussion
1200-1300 Lunch

TOR #4 Examine potential factors responsible for retrospective patterns.

1300-1430 Report of Working Group on Retrospective Patterns in VPA
Working Paper 4.1 - (Legault)
1430-1515 Discussion
1515-1530 Break

TOR #3 Model implications of zeros in fishery independent indices.

1530-1600 *Working Paper 3.1* - (Legault)
1600-1620 *Working Paper 3.2* - (Terceiro)
1620-1640 *Working Paper 3.3* - Report of ICES methods working group (Legault)
1640-1730 Discussion
1730-1800 Summary /Followup (Chair)
1800 Adjourn

Tuesday February 26

0900-0920 Progress review and Order of the Day (Chair)

TOR #2 Compare utility of statistical catch-at-age models

0920-0940 Overview of SCAA approaches—(Jacobson)
0940-1010 *Working Paper 2.1a* – Georges Bank Yellowtail (Jacobson)
1010-1030 Discussion
1030-1045 Break
1045-1115 *Working Paper 2.1b* - White Hake (Sosebee)
1115-1130 Discussion
1130-1200 *Working Paper 2.1c* - Georges Bank Cod (O'Brien)
1200-1230 Discussion
1230-1330 Lunch
1330-1445 *Working Papers 2.2, 2.5* - ASPM model - Georges Bank Yellowtail and Gulf of Maine Cod – (Butterworth)
1445-1530 Discussion
1530-1545 Break
1545-1600 *Supplementary paper* and discussion on domed selectivity TOR 2. (Hart/Miller)

1600-1630 *Working Paper 2.3* - Gulf of Maine Cod –ASAP model (Shepherd)
1630-1745 Discussion—GOM cod
1745-1800 Summary/Followup (Chair)
1800 Adjourn

Wednesday February 27
0900-0930 Progress review and Order of the Day (Chair)
TOR #2 Compare utility of statistical catch-at-age models (cont.)
0930-1000 *Working Paper 2.4* - Comparative Simulation Tests—Overview (Legault/Brooks)
1000-1045 *Working Paper 2.4a* **–** Retrospective Pattern: GB Yellowtail (Legault)
1045-1100 Break
1100-1130 Discussion—GB Yellowtail
1130-1200 *Working Paper 2.4b* – Ageing Error: White Hake (Legault/Brooks)
1200-1230 Discussion—White Hake
1230-1330 Lunch
1330-1430 *Working Paper 2.4c* - Domed Selectivity: Georges Bank Cod (Legault/Brooks)
1430-1530 Discussion—GB cod
1530-1545 Break
1545-1730 General Review—SCAA/Simulation test
1730-1800 Summary/Followup (Chair)
1800 Adjourn

Thursday February 28
0900-0920 Progress review and Order of the Day (Chair)
TOR #5 Recommendations on Model Selection for each stock
TOR #6 Linkage to Biological Reference Points
0920-1000 Working *Papers 5.1, 6.1* - Model Recommendations/ Selection Criteria (Rago and Population Dynamics Branch)
1000-1045 Reviews by Species
1045-1100 Break
1100-1230 Reviews by Species (cont)
1230-1330 Lunch
1330-1430 Reviews by Species (cont)
1430-1530 Ecosystem Models for Reference Points—Progress Update (Fogarty/Link/Overholtz)
1530-1730 Revisit Topics as Needed
1730-1800 Synthesis and Report Planning (Chair)
1800 Adjourn

Friday February 29
0900-0930 Progress review and Order of the Day (Chair)
0915-1030 Follow-up Sessions if Necessary
1030-1045 Break
1045-1230 Report development
1230-1330 Lunch
1330-1600 Summary and Assignments
1600 Adjourn

Appendix 4. List of Working Papers for the GARM-III Assessment Methodology "Models" Meeting (25-29 Feb. 2008)

1.1 Rago, et al. 2008. Data Summary for Nineteen Groundfish Stocks in the Northeast U.S.
1.2 Rago, et al. 2008. Overview of Assessment Methods and Model Selection Criteria for Nineteen Groundfish Stocks in the Northeast U.S.
M.1 Mayo R, Col L, Traver M. 2008. Data Summary of Catch and Abundance Measures. (for Working Papers 1.1 & 1.2)
1.3 Hendrickson L, Col L. 2008. Maps Showing NEFOP Sampling Coverage and Management Areas
2.1 Jacobson L, Legault C, O'Brien L, Sosebee K. 2008. Utility of Statistical Catch-at-age Models for Assessing Northeast Groundfish Stocks (a workshop report)
2.2a Butterworth DS, Rademeyer RA. 2008. Statistical Catch-at-age Analysis *vs* ADAPT-VPA: The Case of Gulf of Maine Cod
2.2b Rademeyer RA, Butterworth DS. 2008. Retrospective Analysis for the Gulf of Maine Cod ASPM Reference Case Assessment
2.3 Shepherd G. 2008. Comparison of ADAPT VPA and ASAP Models for Gulf of Maine Cod
2.4 Brooks L, Legault C, Nitschke P, O'Brien L, Sosebee K, Rago P, Seaver A. 2008. Evaluation of NMFS Toolbox Assessment Models on Simulated Groundfish Data Sets
2.5 Butterworth DS, Rademeyer RA. 2008. Application of and Age-Structured Production Model to the Georges Bank Yellowtail Flounder
3.1 Legault C, Seaver A. 2008. Simulation Studies of Issues Associated with Filling Zeros in VPA Tuning Indices
3.2 Terceiro M. 2008. The Treatment of "Zero" Observations in the Summer Flounder ADAPT VPA Calibration
3.3 ICES Resource Management Committee. 2007. Report of the Working Group on Methods of Fish Stock Assessment (WGMG)
4.1 Legault C, et al. 2008. Report of the Retrospective Working Group
5.1 Rago P, et al. 2008. Recommended Modeling Approaches by Stock Initial Proposals for Consideration by the GARM III Review Panel
6.1 Rago P, et al. 2008. Sufficiency of Models for Biological Reference Points

Supplementary Papers

TOR 1 Gulf of Maine Winter Flounder SCALE Run 2
TOR 2 Miller T, Hart D, Cadrin S, Jacobson L, Legault C, Rago P. 2008. Analyses of Tagging Data for Evidence of Decreased Fishing Mortality for Large Yellowtail Flounder
TOR 2 Butterworth DS, Rademeyer RA. 2008. On Drawing Inferences Concerning Trends in Selectivity with Age from Tag-Recapture Information
TOR 2 Jacobson L. 2008. Questions about the Adjusted Lognormal Error Distribution used in Calculating Goodness of Fit for Survey and Commercial Age Composition Data in Preliminary ASPM Models.
TOR 2 Jacobson L. 2008. Pope vs Baranov

Appendix 5. List of Participants GARM III Models Meeting February 25-29, 2008

Name	Affiliation	Email
Bob O'Boyle	Beta Scientific (GARM Chair)	betasci@eastlink.ca
Paul Rago	NMFS	paul.rago@noaa.gov
Yan Jiao	Virginia Tech	yjiao@vt.edu
Paul Medley	CIE Reviewer	paulahmedley@
Toni Chute	NMFS	toni.chute@noaa.gov
José DeOliveira	CEFAS Lowestoft Lab	jose.deoliveira@cefas.co.uk
Cynthia M Jones	Old Dominion University	cjones@odu.edu
James Ianelli	AFSC	jim.ianelli@noaa.gov
Stratis Gavaris	Fisheries & Oceans Canada	gavaris@mar.dro-mpo.gc.ca
Fred Serchuk	NEFSC, Woods Hole	fred.serchuk@noaa.gov
Jim Weinberg	NEFSC, Woods Hole	james.weinberg@noaa.gov
Maggie Raymond	Associated Fisheries of Maine	maggieraymond@comast.net
Tom Nies	New England Fisheries Management Council	tnies@nefmc.org
Rich McBride	NEFSC, Woods Hole	richard.mcbride@noaa.gov
Rebecca Rademeyer	University of Cape Town	rebecca.rademeyer@uct.ac.za
Doug Butterworth	Univesity of Cape Town	doug.butterworth@uct.ac.za
Tim Miller	Pop Dy, NEFSC	timothy.j.miller@noaa.gov
Liz Brooks	Pop Dy, NEFSC	liz.brooks@noaa.gov
Paul Nitschke	Pop Dy, NEFSC	paul.nitschke@noaa.gov
Susan Wigley	Pop Dy, NEFSC	susan.wigley@noaa.gov
Larry Alade	Pop Dy, NEFSC	larry.alade@noaa.gov
Steven Correia	Mass DMF	steven.correia@state.ma.us
Chris Legault	NEFSC, Woods Hole	chris.legault@noaa.gov
Kohl Kanwit	ME DMR	kohl.kanwit@maine.gov
Mike Palmer	Pop Dy, NEFSC	michael.palmer@noaa.gov
Laurel Col	Pop Dy, NEFSC	laurel.col@noaa.gov
Jessica Blaylock	Pop Dy, NEFSC	jessica.blaylock@noaa.gov
Gary Shepherd	Pop Dy, NEFSC	gary.shepherd@noaa.gov
Yanjun Wang	St. Andrews Biological Station, D 70	wangy@mar.dfo-mpo.gc.ca
Lou Van Eeckhaute	Fisheries & Oceans Canada	van-eeckhautel@mar.dfo-mpo.gc.ca
Mark Wuenschel	NEFSC, Woods Hole	mark.wuenschel@noaa.gov
Andrea Strout	NEFSC, Woods Hole	andrea.strout@noaa.gov
Frank Almeida	NEFSC, Woods Hole	frank.almeida@noaa.gov
Loretta O'Brien	Pop Dy, NEFSC	loretta.o'brien@noaa.gov
Ralph Mayo	Pop Dy, NEFSC	ralph.mayo@noaa.gov
Mark Terceiro	Pop Dy, NEFSC	mark.terceiro@noaa.gov
Katherine Sosebee	Pop Dy, NEFSC	katharine.sosebee@noaa.gov
Lisa Hendrickson	Pop Dy, NEFSC	lisa.hendrickson@noaa.gov
Tom Wood	NRC/DMAST	awood@gso.uri.edu
Dan Goethel	SMAST	dgoethel@umassd.edu
Steve Cadrin	NOAA/Umass CMER	steven.cadrin@noaa.gov
Bill Overholtz	NMFS	william.overholtz@noaa.gov
Jiashen Tang	NEFSC, Woods Hole	jjtang@mercury.wh.whoi
Larry Jacobsen	NEFSC, Woods Hole	larry.jacobsen@noaa.gov
Dvora Hart	NEFSC, Woods Hole	dhart@mercury.wh.whoi

Appendix 6. Statement of Work of Center for Independent Experts at GARM-III "Models" Meeting

General

The Groundfish Assessment Review Meeting (GARM) brings together stock assessment experts to peer review work on the status of 19 important fish stocks that are managed by the New England Fishery Management Council. GARM-III takes place in 2007-2008, and it will consist of four meetings that are cumulative in nature (i.e., successive meetings incorporate methods and results that were accepted at previous GARM-III meetings). Each meeting will have a chairman as well as external panelists. A brief description and dates of the four GARM-III meetings are given below:

1. **"Data" Meeting (October 29 – November 2, 2007)**
Review the commercial and survey data that will be used in the stock assessments. Identify appropriate statistical methods for analyzing those data (including bycatch and discard issues, changes in growth rates and other life history traits, issues related to merging databases, etc.). Other sources of data to be considered are tagging programs for cod and yellowtail flounder, and Industry-Based Surveys. Candidate sources of data relevant to ecological and ecosystem considerations will also be described.

2. **"Modeling" Meeting (Feb. 25 – 29, 2008)**
Determine the most appropriate stock assessment methods and models for each of the 19 stocks. Perform runs of those models to obtain results (historical and current estimates of F and B) based on commercial and survey data, probably through calendar year (CY) 2006. Evaluate retrospective patterns and their importance for status determination.

3. **"Biological Reference Point (BRP)" Meeting (April 28 – May 2, 2008)**
Update or redefine BRPs for each of the 19 stocks. Use data available through CY2006. Consider whether the BRPs are reasonable in light of results from the "Modeling" Meeting. Define the appropriate initial conditions for forecasting and rebuilding strategies, particularly with respect to trends in biological attributes, recruitment and survival rates. Comment on relevant ecosystem considerations as they relate to rebuilding strategies.

4. **GARM-III "Final" Meeting (August 4-8, 2008)**
Use all of the methods proposed from the previous three meetings, along with survey and catch information through CY2007, to estimate fishing mortality rates and biomass for each stock. Based on procedures from the BRP Meeting, finalize the BRPs, appropriate initial conditions, and biological assumptions related to forecasts. Determine the status of each stock.

This SOW applies specifically to the GARM-III "Modeling" Meeting, which will take place at the Woods Hole Laboratory of the Northeast Fisheries Science Center (NEFSC) in Woods Hole, Massachusetts, from February 25 -29, 2008. The meeting will have a chairman (non-CIE) as well as external panelists, three of whom will be from the Center of Independent Experts (CIE).

Overview of CIE Peer Review Process

The Office of Science and Technology implements measures to strengthen the National Marine Fisheries Service's (NMFS) Science Quality Assurance Program (SQAP) to ensure the best available high quality science for fisheries management. For this reason, the NMFS Office of Science and Technology coordinates and manages a contract for obtaining external expertise through the Center for Independent

Experts (CIE) to conduct independent peer reviews of stock assessments and various scientific research projects. The primary objective of the CIE peer review is to provide an impartial review, evaluation, and recommendations in accordance to the Statement of Work (SoW), including the Terms of Reference (ToR) herein, to ensure the best available science is utilized for the National Marine Fisheries Service management decisions.

The NMFS Office of Science and Technology serves as the liaison with the NMFS Project Contact to establish the SoW which includes the expertise requirements, ToR, statement of tasks for the CIE reviewers, and description of deliverable milestones with dates. The CIE, comprised of a Coordination Team and Steering Committee, reviews the SoW to ensure it meets the CIE standards and selects the most qualified CIE reviewers according to the expertise requirements in the SoW. The CIE selection process also requires that CIE reviewers can conduct an impartial and unbiased peer review without the influence from government managers, the fishing industry, or any other interest group resulting in conflict of interest concerns. Each CIE reviewer is required by the CIE selection process to complete a Lack of Conflict of Interest Statement ensuring no advocacy or funding concerns exist that may adversely affect the perception of impartiality of the CIE peer review. The CIE reviewers conduct the peer review, often participating as a member in a panel review or as a desk review, in accordance with the ToR producing a CIE independent peer review report as a deliverable. The Office of Science and Technology serves as the COTR for the CIE contract with the responsibilities to review and approve the deliverables for compliance with the SoW and ToR. When the deliverables are approved by the COTR, the Office of Science and Technology has the responsibility for the distribution of the CIE reports to the Project Contact. Further details on the CIE Peer Review Process are provided at http://www.rsmas.miami.edu/groups/cie/

Requirements for CIE Reviewers

Three CIE reviewers are requested to conduct an impartial and independent peer review in accordance with the Terms of Reference (ToR) herein. Each CIE reviewer's duties shall not exceed a maximum of 14 days conducting pre-review preparations with document review, participation on the SARC panel review meeting, editorial assistance for the SARC Chair, and completion of the CIE independent peer review report in accordance with the ToR and Schedule of Milestones and Deliverables. CIE reviewers shall have working knowledge and recent experience in the application of modern fishery stock assessment models. Expertise should include both the use of statistical catch-at-age and traditional VPA approaches. Experience with comparative studies of these approaches is especially valuable. Reviewers should also have experience in evaluating measures of model fit, identifiability, uncertainty, and forecasting. Some experience with groundfish (such as cod, haddock, flounder) population dynamics would be useful.

Specific Activities and Responsibilities

The CIE's deliverables shall be provided according to the schedule of milestones listed on page 5. The GARM Chairman will use contributions from the CIE panelists as well as from other external panelists, to produce the GARM Panel Summary Report. In addition, each CIE panelist will write an individual independent report. These reports will provide peer-review information for a presentation to be made by NOAA Fisheries at meetings of the New England and Mid-Atlantic Fishery Management Councils in 2008. The GARM Panel Summary Report shall be an accurate and fair representation of the GARM panel viewpoint on the quality and soundness of the science, methods and data with regard to each Term of Reference (see Annex 1). The report shall also contain recommendations for improvement that might be implemented in a future GARM meeting.

Charge to GARM panel

The panel is to determine and write down its viewpoint on the quality and soundness of the science, methods and data with regard to each Term of Reference (see Annex 1). Criteria to consider include whether: (1) the data are adequate and were used properly; (2) the analyses and models were appropriate and correctly accomplished; and (3) the conclusions are correct/reasonable. Where possible, the chair shall identify or facilitate agreement among the panelists regarding each Term of Reference.

During the course of the review, the panel is allowed limited flexibility to deviate from the results and recommendations of earlier GARM-III meetings. This flexibility may include minor alterations in procedures previously established at the peer review of the Data Methods Meeting in October 2007. Large scale changes, such as changing a stock definition would not be possible in view of the difficulties of implementing these changes in time available before the final GARM meeting in August 2008.

Furthermore, if the panel rejects certain assessment models, the panel should explain why those particular models are not suitable, and the panel should recommend suitable alternatives. If such alternatives cannot be identified, then the panel should indicate that the existing (status quo) models are the best available at this time.

Roles and responsibilities

Prior to the meeting (GARM chair and CIE panelists)

Review the reports produced by the Working Groups, and read background reports.

During the Open meeting

(GARM chair)

Act as chairperson, where duties include control of the meeting, coordination, control, and facilitation of the presentations and discussions, and ensuring that all Terms of Reference of the GARM are reviewed and completely addressed.

During the question and answer periods, provide appropriate feedback to the assessment scientists on the sufficiency of the analyses and when possible, suggest improved approaches. It is permissible to discuss the working papers, and to request additional information to clarify or revise existing analyses, if that information can be produced rather quickly.

(CIE panelists)

For each model approach, participate in panel discussions on the quality and soundness of the science, methods and data with regard to each Term of Reference (see Annex 1).

During the question and answer periods, provide appropriate feedback to the assessment scientists on the sufficiency of the analyses. It is permissible to request additional information if it is needed to clarify or revise existing analyses, if that information can be produced rather quickly.

After the Open meeting

(GARM CIE panelists)

Each panelist shall prepare an Independent CIE Report (see Annex 2). This report should comment on the quality and soundness of the science, methods, and data with regard to each Term of Reference.

If any modeling approaches are considered inappropriate, the Independent CIE Report should include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report should indicate that the existing modeling approaches are the best available at this time.

During the meeting, additional questions that are not in the Terms of Reference but which are directly related to the assessments may be raised. Comments on these questions should be included in a separate section at the end of the Independent CIE Report prepared by each panelist.

If a panelist feels that his/her comments are adequately expressed in the GARM Panel Summary Report, it will not be necessary to repeat the same comments in the Independent CIE Report. In this case, the Independent CIE Report can be used to provide greater detail on specific Terms of Reference or additional questions raised during the meeting.

(GARM chair)

The GARM chair shall prepare a document summarizing the background of the work to be conducted as part of the review process, and summarizing whether the process was adequate to successfully address the Terms of Reference. If appropriate, the chair will include suggestions on how to improve the process. This document will constitute the introduction to the GARM Panel Summary Report.

(GARM chair, CIE and non-CIE panelists)

The GARM Chair will take the lead in preparing, editing, and completing the GARM Panel Summary Report, based on contributions from the external panelists (CIE and non-CIE). The panelists and the chair will discuss their views on each Term of Reference and whether their opinions can be summarized into a single conclusion for all—or only for some—of the Terms of Reference. For TORs where a consensus view can be reached, the GARM Panel Summary Report will contain a summary of such views. In cases where multiple and/or differing views exist on a given Term of Reference, the GARM Panel Summary Report will note that there was no agreement and will specify—in a summary manner—what the various opinions are and the reason(s) for the different opinions.

The chair's objective during this Summary Report development process will be to identify or facilitate the finding of an agreement, rather than forcing the panel to reach an agreement if this is not possible.

The GARM Panel Summary Report (please see Annex 3 for information on contents) should comment on the quality and soundness of the science, methods, and data with regard to each Term of Reference.

If any modeling approaches are considered inappropriate, the GARM Panel Summary Report should include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report should indicate that the existing modeling approaches are the best available at this time.

The contents of the draft GARM Panel Summary Report will be approved by the CIE panelists by the end of the Summary Report development process. The GARM chair will finalize all editorial and formatting changes prior to approval of the contents of the draft GARM Panel Summary Report by the CIE panelists. The GARM chair will then submit the approved GARM Panel Summary Report to the NEFSC contact (i.e., SAW Chairman).

Schedule of Milestones and Deliverables

The milestones and schedule are summarized in the table below. No later than March 14, 2008, the CIE panelists should submit their Independent CIE Reports to the CIE for review[3]. The Independent Reports shall be addressed to "University of Miami Independent System for Peer Review," and sent to Dr. David Sampson, via e-mail to David.Sampson@oregonstate.edu and to Mr. Manoj Shivlani via e-mail to mshivlani@rsmas.miami.edu

Milestone	**Date**
Open workshop at Northeast Fisheries Science Center (NEFSC)	Feb. 25 – 29, 2008

[3] All reports will undergo an internal CIE review before they are considered final.

(begin writing reports, as soon as open Workshop ends)	
GARM Chair and CIE panelists work at the NEFSC drafting reports	Feb. 28 – 29
Draft of GARM Panel Summary Report, reviewed by all CIE panelists, due to the GARM Chair **	March 14
CIE panelists submit Independent CIE Reports to CIE for approval	March 14
GARM Chair sends Final GARM Panel Summary Report, approved by CIE panelists, to NEFSC contact (i.e., SAW Chairman)	March 21
CIE provides reviewed Independent CIE Reports to NMFS COTR for approval	March 28
COTR notifies CIE of approval of reviewed Independent CIE Reports	April 4 *
COTR provides final Independent CIE Reports to NEFSC contact	April 4

* Assuming no revisions are required of the reports.
** The GARM Panel Summary Report will not be submitted, reviewed, or approved by the CIE.

The SAW Chairman will assist the GARM chair prior to, during, and after the meeting in ensuring that documents are distributed in a timely fashion. NEFSC staff and the SAW Chairman will make the final GARM Panel Summary Report and Independent CIE Reports available to the public. Staff and the SAW Chairman will also be responsible for production and publication of the collective Working Group papers.

Acceptance of Deliverables
By 28 March 2008, CIE shall complete and submit the independent CIE peer review reports in accordance with the ToR, which shall be formatted as specified in Annex 2. Upon review and acceptance of the CIE reports by the CIE Coordination and Steering Committees, CIE shall send via e-mail the CIE reports to the COTRs (William Michaels William.Michaels@noaa.gov and Stephen K. Brown Stephen.K.Brown@noaa.gov) at the NMFS Office of Science and Technology by the date in the Schedule of Milestones and Deliverables. The COTRs will review the CIE reports to ensure compliance with the SoW and ToR herein, and have the responsibility of approval and acceptance of the deliverables. Upon notification of acceptance, CIE shall send via e-mail the final CIE report in *.PDF format to the COTRs. The COTRs at the Office of Science and Technology have the responsibility for the distribution of the final CIE reports to the Project Contacts.

Key Personnel
Contracting Officer's Technical Representative (COTR):

William Michaels
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-713-2363 ext 136

Stephen K. Brown
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
Stephen.K.Brown@noaa.gov Phone: 301-713-2363 ext 133

Contractor Contacts:

Manoj Shivlani, CIE Program Manager
University of Miami, RSMAS, 4600 Rickenbacker Causeway, Miami, FL 33149
mshivlani@rsmas.miami.edu Phone: 305-421-4608

Roger Peretti, NTVI Regional Manager
Northern Taiga Ventures, Inc., 814 W. Diamond Ave., Ste. 250, Gaithersburg, MD 20878
rperetti@ntvifed.com Phone: 301-212-4187

Project Contact:

James Weinberg, NEFSC Contact person and SAW Chairman
NMFS Northeast Fisheries Science Center, 166 Water Street, Woods Hole, MA 02543
James.Weinberg@noaa.gov Phone: 508-495-2352

Request for Changes

Requests for changes shall be submitted to the Contracting Officer at least 15 working days prior to making any permanent substitutions. The Contracting Officer will notify the Contractor within 10 working days after receipt of all required information of the decision on substitutions. The contract will be modified to reflect any approved changes. The Terms of Reference (ToR) and list of pre-review documents herein may be updated without contract modification as long as the role and ability of the CIE reviewers to complete the SoW deliverable in accordance with the ToR are not adversely impacted.

ANNEX 1: Draft Terms of Reference for the GARM-III "Models" Meeting

(Last Revised: Oct. 31, 2007; A final draft will be distributed to the Panel prior to the meeting.)

1. For each stock, consider the applicability of one or more of the following modeling approaches to assess stock status:
   - Index methods
   - Production Models
   - Age- or Length-based Models

2. For certain stocks that are aged, compare and contrast the utility of statistical catch-at-age vs. VPA based models with respect to the following criteria:
   - Retrospective patterns
   - Flexibility to account for alternative parameterizations
   - Ability to incorporate external sources of information, especially tagging and environmental data
   - Ability to estimate parameters incorporating prior, external information.

3. Address the implications of zeros in the evaluation of fishery independent indices.

4. Examine potential factors responsible for retrospective patterns.

5. For each stock, define the assessment model that will be used to determine stock status and productivity characteristics until the next "benchmark" assessment is conducted. Where possible, apply the models to data (probably through 2006), to obtain current and historical estimates of F and B and estimates of uncertainty.

6. Evaluate the sufficiency of the assessment models to estimate measures of stock status consistent with Biological Reference Points.

ANNEX 2: Contents of GARM-III Independent CIE Reports

1. The Independent CIE Report should comment on the quality and soundness of the science, methods and data with regard to each Term of Reference. CIE panelists should consider whether the work provides a scientifically credible basis for developing fishery management advice. Scientific criteria to consider include: whether the data were adequate and used properly, the analyses and models were carried out correctly, and the conclusions are correct/reasonable.

If a panelist feels that his/her comments are adequately expressed in the GARM Panel Summary Report, it will not be necessary to repeat the same comments in the Independent CIE Report. In that case, the Independent CIE Report can be used to provide greater detail on specific Terms of Reference or additional questions raised during the meeting.

2. If any modeling approaches are considered inappropriate, the Independent CIE Report should include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report should indicate that the existing modeling approaches are the best available at this time.

3. Any independent analyses conducted by the CIE panelists as part of their responsibilities under this agreement should be incorporated into their Independent CIE Reports. It would also be helpful if the details of those analyses (e.g., computer programs, spreadsheets etc.) were made available to the respective assessment scientists.

4. Additional questions that were not in the Terms of Reference but that are directly related to the assessments. This section should only be included if additional questions were raised during the GARM meeting.

ANNEX 3: Contents of GARM-III Panel Summary Report

1. The first section the report shall consist of an introduction prepared by the GARM chair that will include the background, a review of activities and comments on the appropriateness of the process in reaching the goals of the GARM. The next section will contain comments on the quality and soundness of the science, methods and data with regard to each Term of Reference. The GARM Panel should consider whether the work provides a scientifically credible basis for developing fishery management advice. Scientific criteria to consider include: whether the data were adequate and used properly, the analyses and models were carried out correctly, and the conclusions are correct/reasonable.

If the CIE panelists, the non-CIE panelists and GARM chair do not reach an agreement on a Term of Reference, the report should explain why. It is permissible to express majority as well as minority opinions.

2. If any modeling approaches are considered inappropriate, the GARM Panel Summary Report should include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report should indicate that the existing modeling approaches are the best available at this time.

3. The report shall also include the bibliography of all materials provided during the meeting and any papers cited in the GARM Panel Summary Report, along with a copy of the CIE Statement of Work.

The report shall also include as a separate appendix the Terms of Reference used for the GARM Models Meeting, including any changes to the Terms of Reference or specific topics/issues directly related to the assessments and requiring Panel advice.

Appendix 7. Presentation Highlights and Discussion

This appendix includes the presentation highlights provided by the senior author of each working paper along the rapporteur's notes of the ensuing discussion. In regard to the latter, the emphasis was to capture the main points made. Some rapporteurs used prose while others bullet style. There was only modest editing of these during preparation of this report. Notwithstanding this, the text give a sense of the main topics discussed, areas of agreement, and areas of future work. While these were referred to by the Panel, statements in this Appendix should not be considered the final conclusions of the Panel, which are stated in the body of this report.

Applicability of Models to Assess Stock Status

Rapporteur: Toni Chute

<u>Working Paper 1.1: Rago P, et al. 2008. Data Summary for Nineteen Groundfish Stocks in the Northeast US</u>

Presentation Highlights

The feasibility of any assessment model approach ultimately relies on the quality of underlying data. To that end this report incorporates the full set of data estimation improvements described and endorsed by the review Panel for the GARM Data Methods Meeting (October 29-November 2, 2007). The major revisions include:

- Improved methods for allocating landings to statistical area
- Revised approaches for estimation of total discards
- Improvements to software for estimation of landings at age

This document is a summary of the progress that has been made to date on the preparation of data for the Groundfish Assessment Review Meeting (GARM). The objective of this report is to provide the necessary background on the availability of data for each species, and to inform discussions on model selection. Each chapter includes a set of core tables that illustrate the available commercial landings and discards, recreational landings and discards, and survey indices. For species that are aged, additional information on biological samples and age composition is also provided. Not all of the age samples for 2007 have been processed.

Discard estimates and CVs are based on the SBRM methodology (based on a discard/kept of all species) described at the GARM Data Methods meeting in October 2007. In some instances, individual analysts varied the approach to more accurately account for known interventions that created large regulatory discards. For some species, discards were estimated using hindcast methods. A generalized database has been developed to facilitate the evaluation of alternative estimators. One of the important new aspects of this report is that measures of uncertainty in the discard totals are now estimated.

Recreational catch and landings have been revised by MRFSS and are summarized for 1981 to 2006. Compared to the more popular recreational species (e.g., striped bass) relatively few length frequency samples for groundfish are taken.

Standardized mean abundance and biomass (kg) indices are available for NEFSC spring (1968-2007), NEFSC autumn (1963-2007) and DFO spring (1986-2007) were summarized. The Massachusetts DMF spring and fall bottom trawl surveys (1982-2007) were are also included for stocks that frequent inshore waters. Numbers at age were also estimated for appropriate stocks.

The data summaries are preliminary. In particular, discard estimates may be revised as the individual analysts review the underlying data in more detail. Some of the chapters include more complete, but preliminary, investigations of model performance. Such chapters do not constitute assessments but will be used as a starting point for more comprehensive analyses.

<u>Working Paper 1.3.Hendrickson L and L. Col. 2008. Maps Showing NEFOP Sampling Coverage and Management Areas</u>

Presentation Highlights

Groundfish discards are estimated for the 2008 GARM from trips sampled by fisheries observers from the Northeast Fisheries Observer Program (NEFOP). The spatial distribution of the observed tows from these trips is shown on a series of GIS maps, by gear type during 1986-2006, in relation to statistical reporting areas and the three groundfish closure areas (CAI, CA II, and the Nantucket Lightship CA).

Discussion on Working Papers 1.1 and 1.3

- The important issue for this meeting is to ensure that the assessment models consider the key processes of each stock through considering changes in management and depletion over time, discards, survey variability, and so forth rather than going over all the data. Key objectives include:
    - Comprehensive evaluation of the estimation error in landings, discards, surveys, and catch-at-age.
    - Four stocks will be used as case studies to highlight major components/processes: Georges Bank yellowtail flounder, Georges Bank cod, white hake, and Gulf of Maine cod. Patterns observed for these stocks may be generalized to remaining stocks.
    - Fish sizes have been changing over time too possibly as a result of several factors including fishing, environmental changes, density dependence and natural mortality; these patterns emerge across all stocks.
- Stock structure is defined by statistical areas which are informed by both the topography of the bottom and fishing areas.
    - Not all GB stocks have the same statistical areas associated with them; some GB stocks occupy statistical areas that other GB stocks do not.
    - Stocks are determined by both natural patterns of abundance and management areas. There is some migration between stocks and tagging programs for yellowtail flounder and cod show promise for quantifying not only the rate but the fluxes among populations.
    - The GOM has colder water and is not as productive as compared to GB
    - Survey strata for fishery-independent bottom trawl surveys are chosen to approximate the statistical areas used to quantify commercial landings data. Survey strata are based on additional information including depth and bottom type.
- Distant-water fleets exploited the resources very heavily until the Magnuson act in 1976.
    - Up to 500,000 metric tons a year were removed, much of it prior to the initiation of NMFS surveys. Large reductions in biomass right around the time of the NMFS surveys began.
- The reporting of landings has changed over time
    - Pre 1994 reporting was voluntary and supplemented with interviews, now it is a mandatory logbook system.
    - There are some problems determining the allocation of landings: e.g. where the fish were caught. These problems were addressed at the GARM Data meeting. Allocation to stock areas is less problematic.
    - Dealer records have not changed much, although now they are reported electronically.

    - Observer coverage varies depending on circumstances such as a special access program and species. Overall coverage for groundfish ranges from 2 to 5%.
- Management has become more than just "a strong suggestion" and there have been some significant improvements in some species.
    - In 1994, it was planned to reduce the fishing by 50% in 5-7 years.
    - Closed areas were implemented in 1996, and rolling closures protect the stocks at certain times of year.
    - The closed area boundaries are heavily fished - 75% of haddock are caught within a few miles of the area boundaries.
- Discard estimation has been improved by stratifying the fleets and estimation by a discard to kept ratio.
    - Observer information is key
    - Hindcasting discard estimation is difficult; both observer and survey data are used.
    - With changes in management, discard patterns have changed over time.
    - The average kept weight and average trip duration seem to be the same whether an observer is on board or not.
    - The probability of capture, given the particular fleet, gear and area plus the management effects are used to estimate probable discard.
- Regarding NMFS survey sampling (over 45 years), there is evidence of abundance clustering over time (high Gini numbers).
    - The definition of the "stock area" is what is used when determining whether there is a high Gini index for a certain species, not the whole survey area.
    - In some cases, the high concentration may be due to spawning aggregations.
    - There is no evidence of bias in the survey, but the realized sampling design is not equally effective (i.e., precise) for all species.
- Number of tows within an area are used to calculate the measure of concentration (i.e. Gini index) , but some of the area within strata are not trawlable for a number of reasons
    - However, many untrawlable areas are only temporarily so.
    - Overall though, the survey coverage is generally good; there are not many holes.
    - Looking at areas of survey gear damage could point out where these untrawlable areas are.
    - Untrawled areas, if consistent over time, may not be that important.
- The definition of the plus group has changed over time due to the intensity of fishing, and changes in size of fish at age due to unknown causes.

<u>Working Paper 1.2: Rago et al. 2008. Overview of Assessment Methods and Model Selection Criteria for Nineteen Groundfish Stocks in the Northeast US</u>

Presentation Highlights

This report provides a summary of the assessment approaches that have been used to date for the 19 GARM stocks. Each section gives a brief overview of the methods that have been used to date, an evaluation of their strengths and weaknesses, and the prospects for implementing alternative approaches. This report complements the data and analyses summarized in Working Paper 1.1.

Model selection requires simultaneous consideration of multiple objectives. Alternative models should be technically superior in terms of modern statistical theory, flexibility to handle heterogeneous data and processes, stability of results, and expanded scope of inference. Alternative models must also allow for the provision of scientific advice with respect to reference points. One aspect of particular importance in the Northeast US is stock rebuilding. The model must include a forecasting component or be linked with forecasting program. This is necessary for the evaluation of alternative management strategies. Such evaluations are also required as components of the economic impacts of alternative harvest levels. Finally, coordination of assessment approaches with states and Canada is important for jointly-managed species.

This report addresses the model selection issues for 19 stocks. Each section includes a brief description and history of current approach, its key strengths and weaknesses, and the feasibility of changing assessment models. At the 2005 GARM, only one stock was assessed using a forward projecting age-structured model. For this GARM there is a general tendency to consider the application of forward projection models for several stocks including Georges Bank and Gulf of Maine cod, Georges Bank haddock, all of the yellowtail stocks, and redfish, and white hake. For other species, the current age based assessment models are performing well and may not need further exploration. Several index-based assessments (e.g., white hake, Gulf of Maine haddock) may be upgraded to age or length based models.

Discussion on Working Paper 1.2

- Relative F equation
    - Catch over three years is average of index of abundance
    - Relative F approach is more of a descriptive tool, not to provide an assessment number.
    - Recruitment is not taken into account when using the averaged 3 indices, allow fish a few years to be caught, but the uncertainty in the survey is why there are 3.
    - In the first and last year of a time series the relative F estimate has a different basis (n-1, or 2 if n=3). All abundances are expressed in biomass, weight per tow.
- Placement of surveys in time may affect the way the data are smoothed since the surveys are conducted at different times of the year.
    - Spatial patterns in the survey data have not affected the definition of central tendency.
    - Finer scale changes in the surveys may change the apparent pattern of the fish.
    - Does one go back and look at the survey areas where the aggregated (high GINI number) fish are found to see whether it is in a specific place. This is not an issue in design-based surveys but knowing the variation from tow to tow may be helpful for the model
- Model vs. index-based assessment
    - ability to age all the species was important
    - Some sampling years were not good so a model had to be used to cover the times of poor age data.
    - Model selection needs to be assisted by considering the summary of data is available and what the time spans and ranges of data are
- Dealer data represent a near census of reported landings and have no sampling error in terms of biomass. Allocation of landings to area and age class are subject to the veracity of the VTR data and the intensity of biological sampling. Re area allocation, are the interviews and VTR comparable?
    - The pre1994 interview program was a sampling program facilitated by expert judgment of the port agents.
    - Port agents would impute statistical area from trips, ports for unsampled trips.
    - Error in statistical area for catch is not huge, but the statistical area assignment is less certain than stock area.
- The magnitude and precision of total discards have important implications for model selection.
    - How accurate is the size frequency coming from the observers for the whole stock? Pretty good, but some of the CVs were pretty high for the catch
    - Overall, error in catch-at-age are probably small compared to the indices at age.
- What is the relative magnitude of error from all of the different data sources?
    - Stocks that have some data peculiarities are the ones that have special modeling needs

    - Are there any commonalities in the assessments that do not work? Generally for flatfish and stocks in the Gulf of Maine.
- Need to put some work into defining uncertainty in the catch (i.e., combining errors in commercial and recreational landings with imprecision of discard estimation).
- Replacement yield models
    - Considers replacement yield constant or as constant fraction of extant stock
    - Need to estimate the other parameters (catchability and population growth rate).
    - Need model to fit what is happening in the present, not what happened 20 years ago.
    - Could fit model to different periods of time (e.g. 5, 10, 15 years) in the historical dataset to examine model mis-specification
    - Paper by Glazer (2008) briefly presented and discussed

Potential factors responsible for retrospective patterns

Rapporteur: Susan Wigley

<u>Working Paper 4.1: Legault C et al. 2008. Report of the Retrospective Working Group</u>

Presentation Highlights

This report summarizes a wide range of work related to retrospective patterns in stock assessment, culminating in conclusions and recommendations. A retrospective pattern is a systematic inconsistency among a series of estimates of population size, or related assessment variables, based on increasing periods of data (Mohn 1999). This pattern of change in estimated values can have severe consequences for management of a stock, potentially resulting in depletion of a stock even though the assessments indicate the targets are being met. Retrospective patterns have been observed in some but not all of the stocks in New England, as well as other stocks around the world. Retrospective patterns are not limited to virtual population analysis, having been observed in a wide range of models including statistical catch-at-age models. Instead retrospective patterns are an indication something is inconsistent in the data or model assumptions. However, retrospective patterns are just one diagnostic for stock assessments and lack of a retrospective pattern does not necessarily imply that all is well.

Simulation analyses have demonstrated a number of sources for retrospective patterns, including, missing catch, an increase in natural mortality rate, or a change in survey catchability. The working group examined a number of potential methods to determine the source of a retrospective pattern using simulated data, but was unable to do so. However, the working group found it does appear possible to identify the timing of an intervention which leads to the retrospective pattern in some cases. Similarly, a number of methods were examined to fix retrospective patterns. While the fixes did in fact remove the retrospective pattern, the new assessment was not always closer to the truth than the original assessment, even though the diagnostics of the new model were good. This means that caution must be exercised when applying any fix to an actual assessment to remove the retrospective pattern.

The working group recommends that stock assessment scientists always check for the presence of a retrospective pattern and that a strong retrospective pattern is grounds to reject the assessment model as an indication of stock status or the basis for management advice. The working group also recommended future research to be conducted on the topic to define objective criteria for acceptance of an assessment with retrospective patterns and to determine what type and level of adjustment in management advice is appropriate through management strategy evaluations.

Discussion on Working Paper 4.1

The simulations were useful in illuminating the complexity and importance of identifying the sources and timing of the retrospective patterns as well as indicating where further research could be directed. This research has a broad application because retrospective patterns can occur regardless of the analytic model and occur not only in Northwest Atlantic stock assessments but also in the Northeast Atlantic, Pacific and Southern Ocean stock assessments.

In addition to the sources of retrospective patterns examined through simulation [changes in catch, changes in natural mortality and changes in survey catchability (q)], it was noted that changes in partial recruitment could also cause retrospective patterns.

Informal meta-analyses which examined commonalities among stocks that exhibit retrospective patterns have been conducted; however, no consistent patterns among stocks have emerged with regard to source or timing. It was suggested that the meta-analysis be expanded to examine survey catchabilities among stocks and across areas.

It was also suggested that swept area values, instead of survey catchabilities, be used when investigating this source of retrospective patterns. The use of swept area values may be helpful in the interpretation of survey catchabilities because the values are cast in more realistic terms by translating the survey catchabilities into proportions that are scale-able and this may allow for easier detection of survey trends that could be at odds with the model. The expectation is that the survey values would be on the order of 1 or 2, not 10.

During the discussion, it was noted that while the timing of the change may be known, it may not be possible to determine which portion of the time series is 'true'. It is insufficient to simply split the series to remove the retrospective pattern because while this could remove or diminish the retrospective pattern, it will not necessarily lead to true estimates. To acquire the true value of the estimates, it is necessary to know both the time of the change and the source of the change.

It was pointed out that the simulations conducted to evaluate if retrospective patterns occur by random chance used both realistic noise values (similar to values observed in Northeast stock assessments) as well as unrealistically large noise values. Regardless of the noise level, there was no evidence that retrospective patterns occur by chance alone. It was noted that additional simulation work could be expanded to include correlated noise (e.g. year effects in surveys).

During the discussion of closed areas as a source of retrospective patterns, it was noted that closed area affects could be from regulatory closed areas or from a concentration of effort in a particular area while survey indices were derived from the entire area. It was emphasized that the source of retrospective patterns created by closed areas was limited to sessile animals because the simulations revealed that the retrospective patterns diminished when movement between the closed and open areas was added into the simulations.

Since retrospective patterns arise from a change within a time series, one possible solution to remove a retrospective pattern would be to shorten the assessment time frame. It was noted that shortening the time series used in the assessment is contrary to the belief that a longer time series is preferable; however, given a retrospective pattern, a shorter time series may provide more accurate estimates of current stock status. Using moving window was also discussed as an alternative to shortening the assessment time series. A moving window may also be beneficial in identifying the timing of the change.

When a retrospective pattern occurs, what action should be taken: use another model, shorten time period, or reject the assessment? No conclusion was reached.

Alternative states of nature are analogous to calculating the confidence intervals around estimates and provide estimate bounds for managers. It was emphasized that to properly determine the alternative states of nature, many analyses would be necessary to evaluate all possible sources when the source and timing of the retrospective pattern was not known.

The panel offered the following suggestions to analysts as they prepare their assessments:

1) check for retro patterns
2) evaluate diagnostics (not only retro patterns) and
3) use other auxiliary information when possible (CPUE, landings, etc.) to identify the source(s) and timing of a change(s) within the time series.

It may be possible to account for differences created by a retrospective pattern; however, it assumes the direction of the difference is known but without the knowledge of the source, there is no mechanism to indicate where the truth lies. Caution should be used.

Management strategy evaluations (MSE) were discussed regarding their ability to appropriately determine adjustment levels when providing management advice. The MSE will be limited because they will be conditioned on a given source and timing of the cause for the retrospective pattern. The MSE would have to be cast in a risk analysis framework in order to evaluate different types of adjustments.

The panel agreed that it is important to communicate the retrospective patterns to managers due to the implications of these patterns on reference points and advice on catch and projections.

Model implications of zeros in fishery independent indices

Rapporteur: Larry Jacobson

<u>Working Paper 3.1: Legault, C. and A. Seaver. 2008. Simulation Studies of Issues Associated with Filling Zeros in VPA Tuning Indices</u>

Presentation Highlights

Surveys occasionally do not encounter fish of a given age in a specific year. Since age-specific survey tuning indices are usually assumed to have a lognormal error, zeros are not allowed. The NEFSC standard procedure is to treat the zeros as missing. However, during the 2006 review of the summer flounder assessment, it was claimed that this procedure introduces a bias into the assessment because no information is being presented to the model when the product of abundance and survey catchability for an age are low. No support was provided for this statement beyond this theoretical argument. An approach was recommended to fill the zeros with 1/6 of the smallest non-zero value in the series. It was claimed that this approach reduced the retrospective pattern in the assessment. This claim was later shown to be false. Many changes were made during the assessment with an end result of higher catch advice. This created a perception that filling the zeros causes higher catch advice and so it was requested that this approach be applied to the Northeast groundfish stocks as well. This working paper presents three studies demonstrating why zeros should not be filled.

The first study simply created a time series of abundance and then created deterministic indices with different detection levels for truncating low values to zero. It was shown that filling a sequence of zeros with any constant value introduces a bias by creating a flat index for these years when in fact the actual population is changing. The second study used a population simulator to create datasets for analysis by virtual population analysis under 4 scenarios: 1) all data is used, 2) tuning indices below a set value were truncated to zero and treated as missing, 3) same as case 2 except the zeros filled with a constant value of 0.01, and 4) same as case 2 except the zeros filled according to the 1/6 approach. Cases 1 and 2 had similar levels of bias while case 4 was highly biased relative to the known underlying population N and F. Case 3 had more bias than cases 1 or 2, but not as much as case 3. These simulations demonstrated no support for the claim that treating zeros as missing introduces bias and also showed that filling zeros with a constant can induce large biases. The third study (presented in the appendix) was a simple regression example to demonstrate why filling zeros with a constant can lead to bias. Random data were created and error added to “observed” values which were truncated to zero below some cut-off. When the zeros were treated as missing a regression of the log of observed and true values had a slope near one with the 90% CI of observations covering zero. When the 1/6 rule was applied, the slope of the regression was strongly bias with the 90% CI not covering zero even though it was nearly twice as large as case 1.

These studies concluded that making up data is bad. In the short term, treating zeros as missing is recommended. In the long term, a different error distribution should be found for use instead of the lognormal which allows for zero values. However, this distribution must allow for the truncation effect instead of just treating all zeros as the same value, as a square root transform would, for example.

Working Paper 3.2: Terceiro, M. 2008. The Treatment of "Zero" Observations in the Summer Flounder ADAPT VPA Calibration

Presentation Highlights

There is no consistent pattern in the identification of the additive constant that minimizes the absolute value of Berry's (1987) *g* statistic. There is no strong relationship between the absolute magnitude of the index values, the length of the time series, the number of zeros, the magnitude of the smallest observed value, or any of the usual statistical moments of the series (mean, maximum, non-zero minimum, CV, skewness, kurtosis), and the value of the additive constant that minimizes *g*. Further, while the "one-sixth" of the minimum observed value was identified as the "best" additive constant in 5 of the 24 (21%) cases examined, this level is not high enough to justify this approach as a reliable rule-of-thumb. In fact, the additive constant of 0.01 was identified as "best" for a higher percentage of series (6 of 24 = 25%). Given the inability to identify a constant that consistently minimizes *g*, the best rule is to maintain the current approach of making no adjustment and continue to treat "zero" observations as "missing."

Working Paper 3.3: ICES Resource Management Committee. 2007. Report of the Working Group on Methods of Fish Stock Assessments (WGMG).

Presentation Highlights

The ICES Working Group on Methods of Fish Stock Assessments (WGMG) met in Woods Hole 13-22 March 2007. During the meeting, a working paper was presented on the topic of zeros in tuning indices. The WGMG discussed the working paper, but did not spend any additional time on the topic. It was noted that the ICES standard approach is to treat zeros as missing values when they occur in tuning indices. The WGMG concluded that a different error structure than the typical lognormal error should be developed which allows the use of zeros. However, simulation testing is required to ensure that such an approach is robust to outliers. The delta approaches were suggested but rejected, while a quasi-likelihood function with quadratic term deserves consideration. The WGMG concluded that one should not change data to fit the model, but rather change the model to fit the data.

Discussion on Working Papers 3.1, 3.2 and 3.3

Comments generally supported the conclusions given in the working papers. It was noted that zeroes in fisheries are more likely a problem of truncation than a problem of true zeroes. Sub-sampling and sample size are important factors in the occurrence of zeroes because, for example, in addition to sampling relatively rare large fish in length composition data, it is necessary to sub-sample relatively rare old ones in the age-length key to observe an old fish in catch-at-age data. A zero survey observation generally means that no fish were taken in a number of tows across a number of strata. The frequency with which zeroes occur may be critical because effects of omitting zeros rarely and at "random" intervals has different consequences than omitting long strings of zero values that occur when very young or old organisms become rare for long periods of time.

It was noted that Berry's (1987) approach was designed for analysis of experimental data, rather than data used in stock assessments. Different approaches (other than logs) to transforming data may be optimal in stock assessment when zeroes are present. The possibility of using the square root transformation (which is applicable to zeroes and tends to standardize variance in lognormal data) was discussed but there was some concern about effects on the scale of the mean and variance as well as the fact that zeros would be treated the same regardless of the true level of depletion leading to their truncation. Over-dispersed Poisson or negative binomial distributions might be useful. It might be useful to vary the composition of the "plus" group for the youngest and oldest ages to eliminate zeroes at these ages. This has implications for the estimation of catchability at age which requires a consistent definition of the age group over time. These suggestions have promise but there was insufficient time to evaluate their merits prior to the GARM.

There was a suggestion to explore robust likelihood approaches to avoid excessive influence on model fitting to data affected by the presence of zeros.

Utility of statistical catch-at-age models

Working Paper 2.1. Jacobson L, C. Legault, L. O'Brien and K. Sosebee. 2008. Utility of statistical catch-at-age models for assessing Northeast groundfish stocks: Overview (a workshop report for sections 2.1a, 2.1b, 2.1c)

Rapporteur: Tim Miller

Presentation Highlights

A training workshop using the SS2 (Stock Synthesis, version 2) statistical catch-at-age stock assessment model for northeast groundfish was held at the Northeast Fisheries Science Center (NEFSC) in Woods Hole, MA during 4-7 February 2005. The workshop was successful in developing potentially useful but preliminary SS2 models for three example stocks. In particular, a model for white hake was developed that spanned a period with missing fishery and survey age data. Previous attempts to assess white hake using a variety of models were problematic. Estimates from an SS2 model for Georges Bank (GB) cod appear relatively precise, presumably because of high quality of the available data. Preliminary SS2 models for GB yellowtail flounder suffered from many of the same problems as VPA models in the last assessment and a number of different model configurations were explored in SS2 that help address these problems.

Experience at the workshop with all three stocks indicates that statistical catch-at-age models are applicable and promising for northeast stocks. SS2 and VPA estimates were generally similar, particularly when configured in similar ways. As exemplified by Georges Bank cod, the data available for northeast stocks are adequate for application of statistical catch-at-age models. In fact, data available for all three example stocks were more comprehensive that the data available for many stocks where statistical catch-at-age models are used routinely. Results for white hake show that statistical catch-at-age models may be useful for stocks with missing age data or stocks currently without an analytical assessment approach. The inherent flexibility of statistical catch-at-age models may be advantageous in dealing with difficult assessment problems and complex data, although results for GB yellowtail flounder indicate that there are no "silver bullets" and that some chronic issues are likely to persist.

Participants at the workshop agreed that SS2 was substantially more difficult to apply than the traditional ADAPT VPA and that training and technical support are absolutely necessary at the outset for new users. Events elsewhere show that even experienced users require technical support and access to the software development team on an ongoing basis.

Although it has been used around the world for a large number of stocks, SS2 was developed originally for relatively slow growing, long-lived and unproductive rockfish stocks on the west coast of North America with limited data and relatively low fishing mortality rates. A number of changes were made to the program to accommodate different circumstances in the northeast. Additional work along these lines will probably be required if SS2 is used for northeast groundfish.

Discussion on Working Paper 2.1

A point was raised that, as with SCAA models, recruitment could be estimated freely with VPA models which supported the presenter's assertion that VPA and SCAA can both be parameterized to model the same processes. Moving to the Baranov catch equation from the Pope approximation was discouraged unless absolutely necessary because of computation burdens especially for Bayesian inferences. Shortening the time step is an alternative to maintaining the Pope approximation but this could increase the number of selectivity parameters. There was some disagreement about whether maximum likelihood estimation of the variance of deviations around the stock recruitment relationship was possible. However, the estimation of this parameter was thought likely to be biased.

One reviewer pointed out that knowledge of the relative weighting of the age composition of the landings and surveys would be helpful in determining the respective influences on the overall objective function in a SCAA model. It was also asserted that the variance of age composition residuals should roughly match the assumed variances. However, the method of dealing with effective sample size of age composition varies among researchers and currently, there is no consensus. Along with variance assumptions for the age composition, distributional assumptions should affect the estimation of parameters in SCAA. Finally, one reviewer suggested using Bayesian approach to help in treating uncertainty in different data components.

<u>Working Paper 2.1. Jacobson L, C. Legault, L. O'Brien and K. Sosebee. 2008. Utility of statistical catch-at-age models for assessing Northeast groundfish stocks: Section a on Georges Bank Yellowtail</u>

Rapporteur: Tim Miller

Discussion on Working Paper 2.1: Section a on Georges Bank Yellowtail

There was only limited discussion on this section of the working paper as much of the time available had been devoted to the overview.

<u>Working Paper 2.1. Jacobson L, C. Legault, L. O'Brien and K. Sosebee. 2008. Utility of statistical catch-at-age models for assessing Northeast groundfish stocks: Section b on White Hake</u>

Rapporteur: Ralph Mayo

Discussion on Working Paper 2.1: Section b on White Hake

Trends in residuals from surveys suggest a change in catchability (q) around 1985. This is likely the result of the doors changes; however there are no conversion factors available from the gear comparison studies.

The fishery selectivity declines sharply after 114 cm. This is likely due to large fish inhabiting deeper water unavailable to some gear types. The drop in survey selectivity may also be due to larger fish inhabiting deeper water out of range of the survey. It is also possible that the fishery may be operating in areas outside of the area covered by the survey.

When asymptotic selectivity was imposed on the fishery, model fit was poor. Therefore, flat top selectivity is not an option conditional on the rest of the model configuration, so additional work is required on the shape of the dome or on adjusting the other parts of the model configuration.

The assessment model generates numbers at age which are then converted to number at length to derive selectivity at length. In years with length compositions but no age compositions, it is possible to use a growth model internally to convert age compositions to length compositions. This conversion can also be done externally using existing methods (e.g Kimura) using, for example, age data from adjacent years. Additional age data may be obtained from observed trips, although these otoliths have not aged.

Based on the plot of length vs. age, the variance of length at age seems large, especially for younger ages. The range of length at age represents the range from both spring and autumn surveys, leading to a wider range than if they were plotted by seasonal age.

Working Paper 2.1. Jacobson L, C. Legault, L. O'Brien and K. Sosebee. 2008. Utility of statistical catch-at-age models for assessing Northeast groundfish stocks: Section c on Georges Bank Cod

Rapporteur: Jessica Blaylock

Discussion on Working Paper 2.1: Section c on Georges Bank Cod

It was remarked that the VPA before this was working adequately, except that the retrospective pattern was a little high or moderate compared to other stocks.

There was discussion on the mimic-VPA model that showed a 180 degree 'flipped' pattern in all three stocks (GB Yellowtail, White Hake, and GB Cod) from the pattern exhibited by the SCAA model. However, further examination revealed that only the GB Cod exhibited this 'flipped' pattern. On this subject, it was pointed out that if the SS2 model had been parameterized to have the same biomass as the VPA, then the retrospective pattern would probably not have flipped between the two models.

On a procedural note, the Chairman reminded the panel that, for all 19 stocks, if the use of a new model is advised, it will be important to be clear about why the 'old model' is not as good as the 'new model'. Why did we change the model? What issues will be addressed better?

Concern was expressed about fitting the likelihood (fixed catch) and its influence on parameter estimation. Fixing the weighting so high (lambda = 100) has a large influence on other parameters in the model. This was done here for both runs because of the assumption the catch was known.

The panel was reminded that these were exploratory runs, but that future sensitivity runs could be done. Suggestions included allowing for a dome-shaped selectivity pattern instead of forcing a flat-topped pattern, and setting selectivity and letting catch vary.

A suggestion was made to let selectivity follow a random walk across years, in the context of parameterizing the SS2 differently to mimic the VPA better. However, caution would be necessary since a change in selectivity implies a change in q.

There was consensus that caution would be necessary if going to a full length-based versus age-based model.

Supplementary Working Paper 2.1: Jacobson L. 2008. Georges Bank Yellowtail; Questions about the adjusted lognormal error distribution used in calculating goodness of fit for survey and commercial age composition data in preliminary ASPM models

Presentation Highlights

The likelihood of age composition data and details of calculation are important in SCCA modeling. For example, in preliminary SS2 models for Georges Bank yellowtail flounder, results for recent years depend on whether the model solution favors (tends to fit) survey and commercial age composition data at the expense of survey trend data, or vice-versa. All approaches are approximate because the statistical distribution of age composition data is hard to specify and model misspecifications complicate the matter further. Still, it is important to understand and weight the advantages of the approach taken, particularly if it is new.

According to equation A21.9 in the working paper 2.2-a by Butterworth and Rademeyer, the negative log likelihood of survey and commercial age composition data in the preliminary ASPM for Gulf of Maine cod is:

$$-\ln\left(L^{CAA}\right)=\sum_{y}\sum_{a}\left\{\ln\left(\sigma_{com}\Big/\sqrt{p_{y,a}}\right)+p_{y,a}\left[\ln\left(p_{y,a}\right)-\ln\left(\hat{p}_{y,a}\right)\right]^{2}\Big/2\sigma_{comm}^{2}\right\} \quad \text{A2.19}$$

Where $p_{y,a}$ and $\hat{p}_{y,a}$ are observed and predicted proportions of the total catch age *a* during year *y*, and the standard deviation σ_{comm} associated with the catch-at-age, which is estimated in the fitting procedure using:

$$\hat{\sigma}_{comm}=\sqrt{\sum_{y}\sum_{a}\left\{p_{y,a}\left[\ln\left(p_{y,a}\right)-\ln\left(\hat{p}_{y,a}\right)\right]^{2}\right\}\Big/\sum_{y}\sum_{a}1} \quad \text{A2.21}$$

Thus, the log likelihood term is calculated assuming that measurement errors in catch-at-age arise from a weighed log-normal distribution. Lognormal distributions are often used for catch-at-age data and not cause for concern. However, the weighting factor (i.e. the observed proportion $p_{y,a}$) for each age is unusual and has uncertain statistical properties that should be evaluated and explained. The text following equation A2.2.1 explains "*Punt (pers. comm.) advised weighting by the observed proportions (as in equation A2.19) so that undue importance is not attached to data based upon a few samples only.*"

The following are three concerns that occur to this reviewer immediately. First, the magnitude of the proportion $p_{y,a}$ has no necessary relationship to the number of samples. A very low observed proportion would naturally occur for a rare event in any number (small or large) of samples. Secondly, this approach gives zero weight to data in age composition bins with zero observations, even though the data likely contain valid and reliable information concerning the rarity of fish of that particular age. Finally, use of the observed, rather than the predicted, proportion implies that the observed proportions are more precise as estimates of the population proportion than the predicted ones. If the observed proportions are more precise, then why bother calculating the predicted values? In any case, the observed values probably include measurement and process errors that contribute variance that is not desirable in weights.

Discussion on Supplementary Working Paper 2.1

This paper was not presented to the meeting due to time constraints and thus there was no related discussion. It was however made available to meeting participants.

<u>Working Paper 2.2a: Butterworth, D.S. and R.A. Rademeyer. 2008. Statistical Catch-at-age Analysis *vs* ADAPT-VPA: The Case of Gulf of Maine Cod</u>

Rapporteur: Bill Overholtz

Presentation Highlights

The WP commences with an historical overview. In 2003, given an estimate of the spawning stock biomass (B^{sp}) in 2001 of only 27% of the corresponding level at MSY (B^{sp}_{MSY}) on the basis of an ADAPT-VPA assessment that used data from 1982 onwards only, the Gulf of Maine cod stock was classified as "overfished" in the context of the Magnusson-Stevens Act, and a recovery plan put in place. However, an alternative Statistical Catch-at-age (SCAA; alternatively termed Age Structured Production Model – ASPM) assessment at the time, which took account of survey data back to 1964, suggested that the stock was above B_{MSY}^{sp}. An independent panel appointed as part of the process to review this and other US Northeast groundfish assessments during that year recommended further investigation of this to better understand the difference. The WP addresses and discusses this issue together with a range of other (sometimes conflicting) suggestions made during a number of reviews of the assessment of this stock over the past decade. It finds that the primary reason for the different results is that the ADAPT-VPA assessment imposed asymptotically flat selectivity-at-age in circumstances where there is strong statistical evidence for dome-shaped selectivity in the data. Making allowance for this under either assessment method reverses perceptions that recent fishing mortalities have exceeded F_{MSY}, and robustly estimates B^{sp} relatively close to B_{MSY}^{sp} rather than below the threshold of 0.5 B_{MSY}^{sp} for an "overfished" ("depleted") classification. Compared to the ADAPT-VPA approach which is limited to the period for which catch-at-age data are available, the SCAA/ASPM approach allows the longer series of research survey data available to be taken into account, thus providing a better basis to estimate management quantities linked to MSY-related targets, and doubling the related precision in some cases. Given that such targets play important roles in the implementation of the Magnusson-Stevens Act, the SCAA/ASPM approach would seem to be preferred over ADAPT-VPA for assessing this stock. The calculations conducted also point more generally to the need for care in treatment of the plus-group in analyses, as well as in use of the Beverton-Holt spawning biomass recruitment relationship which can lead to inappropriately low estimates of B_{MSY}^{sp} in certain circumstances, and to the importance of using flexible parameterizations of selectivity-at-age in SCAA/ASPM assessments to avoid possibly misleading impressions of the precision with which quantities such as natural mortality M can be estimated.

Discussion on Working Paper 2.2a

Discussions which followed are summarized in outline form below:

Statistical Distributions

- Does it matter which distributions one uses? The robustness to alternative statistical distributions can and possibly should be checked.

Ricker vs Beverton and Holt (BH) Stock - recruitment model

- Fitting problem at low stock size.
  - Limited number of observations at high stock size. Is there a decline in recruitment at high stock size? Problem arises if data suggest a negative correlation; BH cannot give a negative correlation; BH will give you a steepness of 1 in these circumstances.
  - Presented formulation provides information on pristine stock size so it helps with the issue of no data at high stock size.
  - If a prior on steepness is assumed, the data are overweighted because of neglect of correlation effects
- Has S/R data from the VPA been considered for guidance? Not yet
- Does the the ASPM have an S/R?
  - It must have since ASPM has production component built in: implies a S/R relationship
  - VPAs, however, are restricted to years with catch-at-age data.
- Parameterize the S/R directly?
  - Bayesian prior, every year there is a deviation from the mean of an S/R.
  - Intermediate world-search routine approximates randomization (random effects) (Maunder and Deriso, 2003)
  - An advantage of internal estimation is that the statistical error assumptions are internally consistent unless the S/R relation is mis-specified.
  - One aspect of mis-specification is the sigma r constraint (p. 52). Should autocorrelation be assumed in the S/R relationship? The current formulation of the model does not include autocorrelation in the residuals. Residuals from S/R are usually not randomly distributed.
- Is Rho fixed?

- Yes run with a fixed Rho first.
- Is there autocorrelation in the S/R data? Errors for several sources. Model cannot resolve, outliers rule the day.
- Are S/R residuals usually autocorrelated (AC)? There is AC in the other data as well, if ignored you are overweighting. At end, this cannot be estimated, resulting confidence intervals are usually too narrow. Though there is often positive AC in residuals, attempts to estimate this in fitting sometimes introduces instability to the estimation.

<u>GOM Cod example</u>

- How about just running with the catches starting in 1893?
  - Was not considered practical because run times would be increase substantially. For true "virgin" biomass, it was suggested that the model would have to start 100-200 yrs earlier.
  - Looked at sensitivities for starting year, things don't change much for management, but carrying capacity (K) change dramatically depending on starting point.
- Concern was expressed that the stock structure of cod had been compromised by excessive fishing on some components, particularly in inshore areas.
  - if true, then it is unlikely that the stock has recovered. Absence of inshore components is likely to impact BRPs.
  - What is tradeoff for including long-past history? Problem is in uncertainty. Long vs short-term uncertainty. Never know the past, but can use it, and sometimes need to use it. Starting in 1893, back to pristine in 1960, catches decreased and stock rebuilt. DeValpine and Hastings (2002) found that a Ricker model often fit better even when the BH was the true model. Highest biomass has poor recruitment. A concern if you have relatively few data points. Need a phase plane plot to show chronology of data? All the high stock sizes are in the 1960s when the stock was being fished down quickly.
- One model run implies that initial biomass in 1893 was at 18-30% of virgin stock size; look at sensitivity runs
  - This value is not considered important because results wash out in first 20 years
  - Most recent years are very similar in most of the sensitivity runs
  - S/R plot only shows 1963 onward, the early points may not be representative of the S/R relationship because high stocks were quickly fished down, i.e., the reproductive potential of the larger stock biomasses may not have been realized.
- Reference points: Parametric or non-parametric S/R? Parametric forms are not mandatory, perhaps this is a case for nonparametric.
  - In the absence of a hypothesized mechanism to justify the overcompensation in a Ricker model, a BH model may be more appropriate.
- Landings: Modern statistics began in the 1930s.
  - 5Y data set starts in the 1940s
  - Data collection systems are different now
  - Perhaps should examine the landings series better and consider the impact of starting other than in 1893
- Likelihood: commercial landings are dominated by ages 5, 6, and 7. The plus group begins at age 7.
  - Using a higher plus group is probably not feasible due to small sample sizes
  - Can all the ages (beyond 7+) be examined for the stock assessment? Yes, this should be looked at. Data suggests a dome for commercial PR.
- Is Pope approximation equation working here? This approximation is known to breakdown at high F.
  - Working paper used Pope because NEFSC was using it at the time

    - Only becomes a problem in 1990, and doesn't seem to present a problem here in this ASPM assessment.
- Is selectivity changing in 2003-2006, it appears so from the bubble plots? Why is selectivity in the survey, which uses a liner, lower for younger ages than the commercial selectivity?
    - The model is based on what the data suggest.
- Why is survey selectivity higher on older fish?
    - Caution needed on two counts: curves are normalized, and correspondence with the data was the primary consideration
- GOM interventions: 1976, 1994, rolling closures, gears.
    - Many different characteristics in the data
    - The ASPM approach lets the data rule the day.
- High reference points for ASPM?
    - Fully selected Fs in ASPM model, not averages over a number of ages as usually quoted for VPA. Compare apples and apples
- Long versus short time series of data
    - Long series good for BRPs, but, may be good idea to use shorter series for status determination.
    - Full time series don't include discards, haddock is a good example, many millions of young fish were discarded pre-1960s. Sensitivities to this should be explored.
- Why do outputs shift back and forth between MLE and MCMC.
    - Acknowledged that results are presented in a mixed mode, Bayesian and frequentist
    - The penalized MLE are also the Bayesian posterior modes. Other Bayesian posterior quantities were not given in every case because of the computational burden, compared to the posterior mode which is much more quickly computed.
- Sensitivity analysis: Management parameters are highly skewed, are priors overly influential?
    - Management parameters are just outputs.
    - The only priors that are informative in the ASPM are on S/R residuals. Uniform prior distributions for S, k, h are all uninformative.
    - Skewness could be real.
- Results: data pre 1982, likelihoods are different for different approaches.
    - Yes, because different data were used. Can't compare two models with different data in AIC sense.
- Clarification: Model used bounds for priors in ADMB.
- MSY calculations: Are S/R functions in R/S units? Intersection point on SR curve, comes from a per recruit perspective. Slope of the replacement line. (The equation relevant to this discussion was subsequently found to have been inadvertently omitted from an appendix in Working Paper 2.2a).
- How much does the choice of a S/R model affect the results? A number of sensitivity tests reported in the paper address this question.)
- Scaling: if you assume a Ricker with poor recruitment in early years, then scale is higher. If all due to catch, then scale is lower. Need to look at sensitivities.

<u>Working Paper 2.3: Shepherd G. 2008. Comparison of ADAPT VPA and ASAP models for Gulf of Maine cod</u>

Rapporteur: Bill Overholtz

Presentation Highlights

Gulf of Maine cod stocks were most recently assessed for GARM II using an ADAPT VPA (Mayo and Col 2006). The same input data was used in a forward projecting catch-at-age model (ASAP) (Legault and Restrepo 1998) for comparison purposes. The ASAP model was setup to evaluate five selectivity scenarios: model 1 selectivity was the average (geometric mean) from the last three years in the VPA; model 2 was a fixed equilibrium selectivity pattern used in the yield per recruit calculations; model 3 was a flat-top pattern fitted to ages 1-5 and held constant (1.0) for ages 6-7; model 4 was a fitted using a single logistic curve; and model 5 was a dome shaped selectivity pattern fitted using a double logistic model.

The ASAP model produced fishing mortality and abundance results very similar to the VPA. Among the configurations examined, the selection curve fitted with a single logistic model had the best diagnostics. However, strong retrospective patterns in the ASAP models suggests further examination is needed to determine if alternative configurations would provide a better model fit.

Discussion on Working Paper 2.3

- Compare retrospective pattern for various model formulations of ASAP.
  - Why are the patterns different? Abundance depends on younger ages, SSB driven by older ages.
- ASAP or other forward projecting models
  - What is the effect of choosing time blocks for selectivity?
  - Model results often sensitive to assumed timing of transitions between blocks.
  - ASAP has no dependence between time blocks. Some models introduce a random walk approach that constrains the rate of change in selectivity between adjacent time steps.
- Likelihood: double logistic had highest likelihood
  - Model 5 had the highest total log likelihood.
  - Does single selectivity pattern apply? Yes
- How many blocks in the ASAP? One block for all models. Fits to catch-at-age must be poor? Yes, but improved for later part of the time series.
- SSB? ASAP had peak in 1989, VPA acts as a smoother. Can’t answer at this point. Rivard wts were used for SSB, Start up approach may account for the differences.
- What was the minimum age of the plus group in ASAP? Age 7.
- How were the means estimated for the 7+ age group? They are weighted averages based on observed values.

<u>Working Paper 2.5: Butterworth, D.S. and R.A. Rademeyer. 2008. Application of an Age-Structured Production Model to the Georges Bank Yellowtail Flounder</u>

Rapporteur: Tim Miller

Presentation Highlights

The WP provides the results for a Reference Case application of ASPM to the Georges Bank yellowtail flounder, together with those for three sensitivities. Strong residual patterns in the fits to survey indices of abundance, particularly to that for the Canadian DFO Spring survey, raise concerns about the compatibility of the population model and these indices. The model fits strongly favor domed over asymptotically flat selectivity. Selectivity assumptions are key to estimates of stock status, with fits to a fully flexible selectivity parameterization indicating the resource to be effectively at its MSY level B^{sp}_{MSY}, whereas the imposition of asymptotically flat selectivity sees the stock estimated to be below the “overfishing” threshold of 0.5 B^{sp}_{MSY}.

Discussion on Working Paper 2.5

There were only points of clarification made with discussion held off until after the follow-up dome-selectivity analysis of tagging data presented by Tim Miller.

Supplementary Working Paper TOR 2: Miller, T, D. Hart, S. Cadrin, L. Jacobson, C. Legault, and P. Rago. 2008. Analyses of tagging data for evidence of decreased fishing mortality for large yellowtail flounder

Rapporteur: Liz Brooks

Presentation Highlights

We performed two analyses of yellowtail tagging data from an ongoing cooperative NMFS tagging study. The first compares expected probability of recovery by age class for tagged fish based on estimates of age-specific fishing mortality by Working Paper with the observed proportions of recoveries (by sex) for different length classes (and approximate corresponding ages) in the yellowtail tagging data. In the second analysis, we fit a series of finite-state continuous-time models to the yellowtail flounder tagging data to estimate different fishing mortality parameters by length class at release while also estimating migration and natural mortality rates along with reporting probability and a scalar to adjust fishing mortality in the first month after release. None of our analyses showed evidence of decreased selectivity for large yellowtail flounder.

Discussion on Supplementary Working Paper TOR 2

- Presentation stated dimorphic mortality could be contributing to differential return rates between males and females
    - It was queried whether there was evidence or published results to this fact. Presenter replied that the survey has skewed sex ratio at older ages, and also pointed to tagging results (Cadrin et al) that suggest lower recovery rate of males—which could be explained by higher natural mortality on males (among other possible explanations).
- Presentation offered clarification regarding the net flux in movement (instantaneous rates) versus the realized difference in abundance between the three regions (given vastly different stock sizes in each area).
- Reviewer asked for the size bins for the 3 size classes—those values were not immediately available.
    - Motivation for specifying the break points was to examine the veracity of the dome proposed in working paper 2.5 rather than to attempt to estimate the true selectivity at size.
- Reviewer asked for clarification on the negative correlation between M and reporting rate; reviewer asked for the difference in high reward value
    - It was clarified that there was a $100 instant winner for high reward, but a regular tag got put into a once-a-year $1000 lottery.
    - A question asking for the difference in reporting rate, assuming high reward equaled 100% reporting and 60% reporting of regular tags
    - It was also stated that the overall recovery rate is 13% (presumably for all tags, both high and regular reward).
- A question was asked regarding detection probability for the tags; it was responded that the fish go through a filleting machine, and that they have gotten returns from the processors, so the tags could be missed at sea but landed fish would most likely have tags detected.
- It was asked whether they were to use this information to recommend information in assessments (regarding scale of F in assessments) or just to inform on PR.
    - Were the data meeting conclusions suggesting that these rates of Fishing Mortality were most precise and should be used to check assessment results?

- What are we supposed to do with the different F at the end from assessment vs. the tagging data?
- Clarification was made that the tagging data were to inform the assessment, and in particular in this case, it was informing as to partial recruitment rather than providing estimates of F.
- Regarding estimated movement (pattern and magnitude), it was noted that the high rate of perceived movement could be a result of where the fish were tagged (i.e. right on the border between SNE and GB).
- Question regarding mixing of one month only after release
  - hypothesis of fully mixed after one month had not yet been tested, but it could be tested by modeling two periods, allowing the first period to vary in length, and then determining when the mixing coefficient in the second period became statistically non-significantly different from 1.0.
- A question was asked about recoveries in closed areas? A conclusion (second bullet) was that no evidence for emigration of yellowtail to un-fished areas (however this bullet related to areas outside the three stock boundaries).
  - This bullet referred to the dome selectivity resulting from a closed area.
- Another question was asked about tag reporting in the case of discards? This would decrease the reporting rate.
  - It noted that they did receive tags from the scallop fleet, and that those recoveries were not only from observed trips.
  - Presenter also noted that only very small amounts of yellowtail flounder were retained in scallop fishery.

<u>Supplementary Reply TOR 2: Butterworth, D.S. and R.A. Rademeyer. 2008. On Drawing Inferences Concerning Trends in Selectivity with Age from Tag-Recapture Information</u>

Rapporteur: Liz Brooks

Presentation Highlights

The WP shows that the implications of dome shaped selectivity for tag recovery proportions as a function of age depend on whether the drop in selectivity at large age arises from a gear selection effect or is surrogating emigration. A simple extraction of summary statistics from tag-recapture data, including a measure of mean tag return time, is suggested to throw further light on the mechanisms actually in play for various stocks, as current high estimates of M from tagging studies may be consequences of failing to take account of emigration to outside the fishing area.

Discussion on Supplementary Reply TOR 2

- Presentation suggested that fish behavior, specifically swimming strength of older fish, could create a dome shape.
  - It was pointed out that if one extended the “strong swimming” argument further, one would expect eventually that selectivity of slow swimmers would continuously decrease availability of slow swimmers until you only have fast swimmers surviving.
- Presentation also suggested that emigration from the fishing grounds would make fish unavailable and could also generate a dome-shaped selectivity.
  - Reviewer asked about the fact that estimated migration rates were low (as a means of arguing against emigration masking selectivity). Presenter suggested that there could be some confounding between reporting rate and an emigration rate. There was debate regarding mean time to recovery and confounding between emigration and reporting rate.

  - It was explained that this has been investigated (mean time between release and recapture) and that no age-specific differences could be found.
  - Presenter countered that mean time to recovery would also be low if emigration occurred.
- It was noted that the Miller et al. model incorporates the time specific info by modeling recaptures individually.
  - While the presenter's mathematics provides an informative check, the Miller model can incorporate these same sorts of features.
- It was interjected that these hypotheses don't sound biologically plausible for yellowtail.
  - Presenter agreed that *a priori* these hypotheses sound far more plausible for cod.
  - One still needs a hypothesis to explain the disappearance from both survey and fishery at 4-5 year olds.
  - It was suggested that a more credible explanation is a higher M on males from age 3 on.

Working Paper 2.4: Brooks, L., C. Legault, P. Nitschke, L. O'Brien, K. Sosebee. P. Rago and A. Seaver A. 2008. Evaluation of NMFS Toolbox Assessment Models on Simulated Groundfish Data Sets.

Rapporteur: Tim Miller

Presentation Highlights

A simulation study was performed to evaluate the performance of five NOAA Fisheries Toolbox assessment models (AIM, ASPIC, SCALE, VPA, and ASAP). Data sets corresponding to three representative groundfish stocks (Georges Bank yellowtail flounder, Georges Bank cod, and white hake) were simulated with PopSim, a simulation program in the Toolbox. For each simulated stock, a base case data set was produced as well as three data sets with a known error. There were 12 data sets in total (three stocks with four data sets each) and for each data set, 100 random realizations were generated with PopSim. Each model performed an "assessment" on the simulated datasets, and the results were compared with the "true" value (i.e., the known parameter values used to generate the data sets). Results for each model in each of the 12 cases were summarized with respect to bias and precision (CV). The base case served as a benchmark to determine how well each model could replicate the truth, and as a point of comparison for model performance on the data sets with known error. In general, no model was a clear winner in all cases. Data sets that reflected errors associated with sampling (aging error or number of length samples) were best handled by models that either did not use age (AIM) or models that incorporate error into catches (ASAP). The VPA, because it matches catch exactly, suffered the most bias and had the poorest precision in these cases. However, when the source of error introduced a "break" in the time series (as in all of the yellowtail flounder cases), none of the model configurations was robust to the effect. The "east coast" approach of tuning to age-specific survey indices appears to be robust to the shape of the selectivity function. In the case of misspecification of the fleet selectivity (assuming logistic when it is dome), both forward and backward projecting models were impacted, but the effect was only apparent at the oldest ages (as would be expected). ASPIC failed in all simulated data sets, but this was due to the nature of the simulated data (all of which were one-way trips), and not to deficiencies in the model.

Discussion on Working Paper 2.4: Overview

There were only a few clarifications on the generalities of the presentation before getting into the specific sections.

Working Paper 2.4: Brooks, L., C. Legault, P. Nitschke, L. O'Brien, K. Sosebee. P. Rago and A. Seaver A. 2008. Evaluation of NMFS Toolbox Assessment Models on Simulated Groundfish Data Sets. Section a on Georges Bank Yellowtail

Rapporteur: Tim Miller

Discussion on Working Paper 2.4: Section a on Georges Bank Yellowtail

In general, ASAP and VPA performed similarly with regard to bias and precision, but ASAP is somewhat more flexible. For Georges Bank Yellowtail, it was pointed out that other work that simulated data over longer time series showed bad fits of ASPIC as well.

One participant wondered whether the differences in the behaviors of the models was related to the weightings of different data components was the cause. It was pointed out that a simpler model like AIM would put zero weight on some data components. One panelist wondered whether changes in the signal could be picked up with the age-structured models, but the presenter pointed out that, while this is true, which data component was the cause would still be unknown. However, one reviewer pointed out that detecting these changes is more difficult with the simpler model AIM, but perhaps this could be modified to do so. It was also pointed out that the behavior of SCALE not being any better than other models was somewhat surprising since selectivity is set up as a function of length in POPSIM.

Working Paper 2.4: Brooks, L., C. Legault, P. Nitschke, L. O'Brien, K. Sosebee. P. Rago and A. Seaver A. 2008. Evaluation of NMFS Toolbox Assessment Models on Simulated Groundfish Data Sets. Section b on White Hake

Rapporteur: Ralph Mayo

Discussion on Working Paper 2.4: Section b on White Hake

Poor performance of AIM relative F estimates is likely to have been a re-scaling problem associated with conversion procedures used to back transform relative F to original units. Ageing error per se should not have an influence on AIM results since it uses pooled estimates of biomass.

Working Paper 2.4: Brooks, L., C. Legault, P. Nitschke, L. O'Brien, K. Sosebee. P. Rago and A. Seaver A. 2008. Evaluation of NMFS Toolbox Assessment Models on Simulated Groundfish Data Sets. Section c on Georges Bank Cod

Rapporteur: Jessica Blaylock

Discussion on Working Paper 2.4: Section c on Georges Bank Cod

Some clarification was provided as to the plus group in this analysis (10), in relation to the dome selectivity pattern.

There was discussion concerning why ASAP did not do better with the fleet domed PR. This was because a single logistic growth was estimated, so the flat-topped pattern was forced. This was accepted to be an important question, and that one should also look at the confidence intervals in this context.

Recommendations on Model Selection for each stock

Working Paper 5.1: Rago P, et al. 2008. Recommended Modeling Approaches by Stock Initial Proposals for Consideration by the GARM III Review Panel.

Rapporteurs: Mike Palmer and Kathy Sosebee

Presentation Highlights

This Working Paper provides a review of initial recommendations of anaylsts for preferred assessment models. The appropriate assessment model for each stock is influenced by attributes of the species and their fisheries, and the ability of the model to capture the salient features of the stock dynamics. A primary consideration is the availability of age data. Before 1995 the selection of models for Northeast groundfish was less complicated because fishing mortality greatly exceeded natural mortality and incoming cohorts were quickly fished out. Various management measures reduced fishing mortality and altered the spatial pattern of fisheries. In particular, large closed areas on Georges Bank differentially affected more sedentary species which would benefit from the reductions in F afforded those fish which remained in closed areas. Increases in abundance for some stocks such as Georges Bank yellowtail flounder and haddock were dramatic. As population age structure broadened, the importance of model features such as plus groups became more important as the number of ages in the plus group increased. Model assumptions that were tenable under high F became less so with reduced F. This has motivated the exploration of a class of forward projecting models that allow greater flexibility for characterizing recent trends and extension of assessments back to periods when landings were much greater.

Individual stock dynamics were summarized with six-panel plots that illustrate the relationships among survey abundance, catches, relative F and replacement ratios. High correlations between the replacement ratio and relative F suggest that the population’s rate of growth is responsive to the rate of removal. In turn this suggests that parametric models may be applicable. Disparities that arise at the end of a time series are also important and may be indicative of process changes that lead to retrospective patterns.

These recommendations herein will be reconsidered during the meeting as the results of alternative models and simulation tests are reviewed by the Review Panel.

Two appendices were included. The first examined statistical properties (mean, CV, Gini indices, and estimated design effect) of the NEFSC fishery-independent bottom trawl surveys. The second provided an overview of the AIM (An Index Method) model and an example application.

<u>Working Paper 6.1: Rago P, et al. 2008. Sufficiency of Assessment Models to Estimate Measures of Stock Status Relevant to Biological Reference Points</u>

Presentation Highlights

The definition of Biological Reference Points (BRP) for fish stock is an essential component of stock assessment. Measures of abundance and harvest rates derived from assessment models are compared to standards that constitute desirable states for each stock. These states are designed to achieve maximum sustainable yield and may include some consideration of uncertainty in estimation. For stocks that are subject to mandatory rebuilding programs, the assessment model must produce outputs that can be forecast under various harvest scenarios. This working paper summarizes the basic approaches for estimation of BRPs associated with the candidate models and highlights special considerations associated with reference point estimation.

Biological reference points can be derived as part of the model identification and estimation process. These can be called internal estimates of BRPs as they rely on specification of stock recruitment relationship within the assessment model. The derived parameters can either be used to directly define reference points or, where analytical solutions are more complicated, to parameterize simulation or forecasting models to derive BRPs and measures of uncertainty. Internal estimates are advantageous since they incorporate the full uncertainty of the model estimation. This can also be a disadvantage when the model does not fit particularly well. In these cases, the BRPs can be unstable, varying with minor changes in model configuration.

"External" estimators of BRPs use model outputs of abundance, SSB, recruits and fishing mortality as inputs to stand-alone models. In the Northeast these include stock recruitment models (SRFIT), yield per recruit models (YPR), and stochastic population projection models (AGEPRO). The SRFIT program uses AIC methods to identify appropriate models from either Beverton-Holt or Ricker stock-recruitment models with and without correlated error terms. When an acceptable model can be defined, standard approaches can be used to estimate Fmsy and Bmsy values.

If none of the parametric models are acceptable, a nonparametric method is used to estimate proxy values for Fmsy and Bmsy. These proxies are derived by combining standard yield per recruit (YPR) and SSB per recruit (SSB/R) methods with model based estimates of absolute recruitment. Model parameters can be used to define appropriate partial recruitment vectors for YPR analyses leading to estimates of Fmax. Fmax serves as a proxy for Fmsy. SSB/R estimates for F=Fmax can be multiplied by some function of the recruitment time series to obtain an estimate of SSBmsy or Bmsy. The term "some function" can imply a simple mean of the recruitment series, a measure of central tendency, or a restricted . Consideration of ecosystem conditions, trends in other populations or evidence of environmental trends can be relevant. A simplified overview of the candidate models used for estimating stock status and their relationship to biological reference points is provided.

Discussion on Working Papers 5.1 and 6.1(stocks ordered as per working paper 6.1)

<u>Georges Bank Cod</u>

The Panel recommended inclusion of historical catch (back to 1930s) and surveys (back to 1963) in the calculation of biological reference points to provide additional information on stock productivity. Presence of the retrospective pattern in the VPA suggests that after 1995, there may have been changes in the fishery selectivity. There were major changes to the management structure (closed areas, mesh size) at this time which may explain these changes. The Panel felt that model formulations should be examined which investigated changes to the selectivity curve, particularly with regards to the older age classes in the recent time period (post 1995). No preference was given to whether changes in selectivity were examined in a VPA or SCAA model.

<u>Georges Bank Haddock</u>

The Panel recommended including catch (landings and discards) back to 1930's referring to the VPA performed on historical data in Clark et al., 1982. It was also recommended that model formulations be explored that consider changes in selectivity over time with a particular focus on the recent time period. A recommendation was made by the Panel to explore a domed shaped selectivity pattern though the specifics were not discussed. It was pointed out that recent changes in size at age will have ramifications on the calculation of the biological reference points. There was discussion on a parametric vs. nonparametric fitting of the stock recruit relationship, noting that a non-parametric approach was chosen the last time biological reference points for this stock were discussed (NEFSC, 2002b) which suggests that this may be required again. This requires exploration.

<u>Georges Bank (GB), Southern New England -Mid-Atlantic (SNEMA) and Cape Cod- Gulf of Maine (CCGoM) Yellowtail Flounder</u>

The Panel did not recommend any changes to the current Major-Change Georges Bank VPA model (split survey q's). This is the same model formulation accepted during the most recent TRAC benchmark review (TRAC, 2005). The Panel supported efforts to build on the current formulation of the model to resolve observed retrospective patterns. The Panel recommended splitting the time series for the

SNEMA and CCGOM stocks to resolve observed retrospective patterns in these assessments. The suggested areas of exploration included (1) investigation of spatial differences in the survey selectivity and/or environmental covariates (temperature) which may explain changes; (2) partial recruitment of the plus-age groups (particularly with regards to the uncertainty in the catch-at-age of the SNE stock); (3) using selectivity blocks; (4) changes in productivity/regimes over time (particularly with regards to severe decline of SNEMA stock). There was some concern that the Major-Change model is not based on an accepted causative process but rather based upon removal of the retrospective pattern. The Panel felt that their responsibility was to recommend which models can be used to offer sound management advice. The Panel made no recommendations on whether explorations should be performed with ASAP or ADAPT which is a software choice based upon ease of use.

Gulf of Maine Cod

The panel discussed whether there was large error in the catch-at-age (CAA). Large recreational catch may result in larger than normal error in CAA. Inclusion of discards may also increase the error due to management actions (i.e. 30 lb trip limit in 1999). However, the coefficients of variation (CVs) on ages 3-5 in the landings are less than 10% while other ages are between 20 and 30%.

There is a weak and inconsistent retrospective pattern for this stock, although the formulation of VPA tends to produce some high fishing mortality (F) on certain ages.

It may be worth changing the assumption on F at oldest age to explore dome-shaped partial recruitment. If a dome appears, a hypothesis is needed to explain the domed selectivity. It may be useful to conduct a risk analysis on the assumption of a dome when flat-topped and vice versa. It may also be useful to examine the tagging data for independent confirmation of a dome and it may be useful to extend the age composition to explore the dome hypothesis.

The Panel found no compelling reason to switch models other than extending the series of catch and survey back in time when reviewing biomass reference points. The landings data prior to the 1930s may not be useful since these data were prorated using stock splits which may or may not have been the most appropriate values. Even with VPA, stock size estimates may be hind-cast using survey data back to 1963 to develop reference points.

Formulation of stock-recruit relationship is important for reference points including whether the relationship is fit internal to the model or externally. The Ricker model provided a good fit but there are may be other issues such as the extreme peak in recruitment (1987 year class). The value for gamma ranged from 0.5 to somewhat over 1. The Panel recommended estimating reference points both externally and internally to examine if there is a difference or if either are appropriate.

Witch Flounder

The Panel recommended exploring the reasons for the retrospective pattern that is seen in this stock. There are no compelling reasons to switch from VPA to SCAA unless this would make it easier to do the exploration.

American Plaice

The Panel discussed the potential problem of combining the data for the Gulf of Maine and Georges Bank portions of the stock. If the relative proportion between the two areas is stable, it is not a problem to combine them. The Panel recommended examining the survey trends for the two areas separately. If the trends are similar, the assessor could continue with a single catch-at-age in either VPA or SCAA. If the trends are different, there is a need to separate the CAA into two areas. Since the discards may not be well-determined, it may be better to move to an SCAA but the separability assumption needs to be explored.

Gulf of Maine Winter Flounder

The Panel recommended exploring the SCALE model since the year-classes which seem to track in the length composition do not appear to track in the age composition. There may be some smearing in the age-length keys. The sex ratio effect on the retrospective pattern was very slight if there were very strong differences between sexes. Something else appears to be causing the pattern. The Panel recommended examining implied fishing effort resulting from the model compared to reported effort in the area. There appear to be conflicting patterns between the recruitment time series and the biomass time series. A risk evaluation may be needed to choose between the consequences of believing one or the other.

Southern New England Winter Flounder

The Panel recommended exploring the reason for the retrospective pattern. It appears that it may be a transient effect that may now be gone. If additional years of data do not make the retrospective pattern reappear, then nothing more is needed. If the pattern reappears, then the assessor should split the survey series at a reasonable time (e.g. 1994). It may be useful to examine if a robust likelihood might remove the retrospective pattern because some of the indices may be very influential as years in the retrospective analysis are dropped.

Georges Bank Winter Flounder

The Panel recommended exploring age-based assessment with ASPIC as fall-back model. It felt that the stock is a candidate for a separable model (SCAA) due to the uncertainty in the CAA during the middle of the time series. However, without first examining results, it was not possible to assess the utility of this approach. The Panel recommended that one approach be chosen and if there are no problems observed then there is no need to explore an alternative approach.

White Hake

The Panel recommended assessing white hake using the SCALE model. It was recognized that there exist problems differentiating between red and white hake in the commercial fishery, particularly with regards to discards. The Panel recommended that the speciation problem should be examined by using survey data to speciate small (< 60 cm) fish and calibrate these results with observer data. There is a potential for sexual dimorphism that may confound a SCALE model and this should be considered. The Panel recommended using all available information as SCALE allows tuning to age data. To ensure that a bridge exists from AIM to SCALE, the Panel recommended running SCALE on the existing data set.

Pollock

The Panel recommended staying with AIM for the current assessment but to provide reasonable justification for selection of reference points. The Panel recommended exploring a state-space production model and examining recruitment variability. The desire for US / Canada scientific collaboration was noted, particularly on the stock assessment.

Redfish

The Panel had questions on the current state of the fishery. It was clarified that there is currently no market. The historical market was driven by a few-centrally owned vessels and once these vessels left the fishery, they were not replaced. Current commercial groundfish gear does not catch sufficient fish to supply a market and there exists no small mesh targeted fishery. Use of an age-structured model was recommended based on strong evidence for infrequent large pulse recruitment which persists for decadal time periods. The major recommendations for model formulation were to relax separable assumptions and use a S / R steepness estimate from west coast species (e.g., Pacific ocean perch), allow σ_R to vary (and go large), but make σ_R small where data are uncertain. The Panel expressed a desire to see an SCAA run performed as a check of the current RED model. An additional suggestion was to investigate an externally derived surplus production model (e.g., Jacobson approach).

Ocean Pout

The Panel recommended exploring other models that consider uncertainty in the relative F (e.g., Glazer, 2008) and/or better incorporate available biological information on the species life history parameters such as a Bayesian biomass dynamics model or catch-at-age model with fixed parameters. It was felt that these approaches could be useful in providing management advice. Sufficiency of available age data upon which to construct a growth relationship to support the later recommendation should be explored.

Gulf of Maine / Georges Bank (GOM/GB) and SNE / MA Windowpane Flounder

Discards are important for both stocks. The Panel recommended exploring the use of SCAA in a Collie-Sissenwine formulation as the assessment model, particularly for the GOM/GB stock as AIM is problematic. For SNE/MA windowpane flounder, AIM is an acceptable model.

Gulf of Maine Haddock

The Panel recommended exploring an age-structured model if time permits. Otherwise, AIM is acceptable. The Panel recommended exploring the spatial distribution of the catch and the surveys to determine if the high values of landings and survey catches may be spillover from the Great South Channel area. This may affect the perception of the productivity of this stock. There appears to be a difference in survey selectivity between fall and spring possibly due to large fish spawning in the spring in areas not sampled by the spring survey.

Halibut

The Panel recommended extending the model to fit to the survey trend and input of a productivity parameter and landings prior to 1893. This probably should be assessed with the Canadian coastal stock of halibut as tagging studies have shown large migrations to Canadian waters.

Appendix 8. Glossary of Terms

Term	Definition
Age-structured Models	Class of assessment models which incorporates age and sometimes size information into the estimation of stock status
Biomass Pool Models	Class of assessment model which does not consider the age structure of a stock but rather describes the relationship between its standing biomass, changes in this due to losses (natural mortality) and additions (recruitment and growth) and yield. The ASPIC software package is commonly used for these models.
Catchability	The parameter (q) relating an index of abundance (e.g. survey) to the stock abundance or I = q N. It is thus a measure of the survey's ability to sample the stock. Generally, the higher the parameter q, the more 'catchable' the fish. The catchability can be either of a group of ages or specific to a defined age e.g. survey catchability at age 5. Typically, changes in q are considered over time while changes in q over age are synonymous with selectivity (a combination of gear vulnerability and availability).
Relative Trends models	Class of assessment model which does not include a parameter for catchability (q) and thus only describes relative as opposed to absolute trends in stock size and fishing mortality. The terms 'index-based' and an 'index method' (AIM) have been used for these models although most assessment models use indices.
Partial Recruitment (PR)	The degree of availability of a stock's age or size group to fishing with one time period on a zero to one scale. Thus a PR of 1.0 for age 5 fish indicates that all age 5 fish are available to being caught – they are 'fully recruited' to the fishery. A PR of 0.5 for age 5 fish indicates that this year – class is only 'partially recruited' to the fishery as only 50% are available to being caught. PR, sometimes called selectivity, is a function of gear vulnerability and availability.
Plus group	Aggregation of data on year classes above a specific age into one grouping. For instance, an age five plus group indicates that the data (e.g. population or catch numbers) on all age five and older fish are aggregated into this group.
Retrospective Pattern	Systematic under - or over – estimation of stock related parameters (typically abundance, biomass and fishing mortality) produced by assessments as additional years of data are added. For instance, an assessment in 2003 might estimate the 2003 fishing mortality to be 0.4 while the assessment in 2008 might estimate the now historical fishing mortality to be 0.8. This under-estimation of fishing mortality by the assessment model, if systematically consistent, is termed a retrospective pattern.
Risk Assessment	Formal process which uses the impact of an hypothesis and the probability of this hypothesis to measure the risk of specified objectives not being met.
Selectivity	The degree of availability of an age or size group of fish to a survey or fishing. When scaled to the group fully selected, this and the partial recruitment are equivalent.
Selectivity blocks	Blocks of years in assessment tables in which the selectivity at age is considered to be the same for all years. Validity of the 'Separable Assumption' is a key feature of the sub-class of assessment models which allows error in the catch-at-age (e.g. SCAA)
Separable assumption	The fishing mortality on an age or size group is considered to be the product of two terms – one age and the other time (typically year) related. The fishing mortality is 'separable' into an age and time effect. This assumption is typically employed in the class of assessment models which allow errors in the catch-at-age (e.g. SCAA).
Statistical Catch-at-	Assessment model which allow errors in the catch-at-age. Using values on stock

age (SCAA)	abundance at some age for each year class, and fishing mortality derived from a separable assumption for specified selectivity blocks, calculates population abundance at age. Comparisons between predicted and observed catch-at-age and abundance indices are made for model fitting. ASAP and SCALE are software packages which allow exploration of this sub-class of model.
Stock - recruitment Relationship (S/R)	Relationship between a stock's spawning biomass and the ensuing recruitment. Two of the most widely assumed S/R relationships are that of Beverton & Holt and Ricker. They differ in the degree of density dependence assumed occurring in the stock, with the latter having a stronger negative relationship between density dependence and stock size.
Virtual Population Analysis (VPA) And Cohort Analysis (CA)	Assessment models which assume negligible error in the catch-at-age. Using values on stock abundance at some age for each year class and the observed catch-at-age, calculates population abundance at age. Comparisons between predicted and observed abundance indices are made for model fitting. VPA and CA differ by the precise form of the calculating catch equation. ADAPT is a software package used to implement VPA.